FINE STRUCTURE ANALYSIS OF THREE GENES CODING FOR PROTEINS OF THE WATER SPLITTING PHOTOSYSTEM II REACTION CENTER

DISSERTATION
zur
Erlangung des Doktorgrades
der Naturwissenschaften
der Universität Düsseldorf
Fachbereich Biologie

vorgelegt von
JULIA MORRIS
aus
St. Louis, Missouri, USA

July 2, 1985

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41543905_0117

FINE STRUCTURE ANALYSIS

OF THREE GENES CODING FOR PROTEINS OF THE

WATER SPLITTING PHOTOSYSTEM II REACTION CENTER

DISSERTATION

zur

Erlangung des Doktorgrades

der Naturwissenschaften

der Universität Düsseldorf

Fachbereich Biologie

vorgelegt von

JULIA MORRIS

aus

St. Louis, Missouri, USA

July 2, 1985

Berichterstatter:

Prof. Dr. R. G. Herrmann
Prof. Dr. C. Hollenberg

Tag der mündlichen Prüfung: July 2, 1985

ABBREVIATIONS

A	Adenine
AP	Ammonium persulfate
Atm	Atmosphere
ATP	Adenosine triphosphate
bp	Base pair
BPB	Bromophenol blue
C	Cytosine
cm	Centimeter
cpm	Counts per minute
CsCl	Cesium chloride
°C	Degree Centigrade
dATP	2'-deoxyadenosine 5'-triphosphate
dCTP	2'-deoxycytidine 5'-triphosphate
dGTP	2'-deoxyguanosine 5'-triphosphate
DMS	Dimethylsulfate
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	2'-deoxynucleotide 5'-triphosphate
DTT	Dithiothreitol
dTTP	2'-deoxythymidine 5'-triphosphate
EDTA	Ethylene diamine tetraacetic acid
EtBr	Ethidium bromide
gm	Gram
g	Gravity
G	Guanine
$\triangle G$	Change in Gibbs free energy
hrs	Hours
kb	Kilobase
kd	Kilodalton
l	Liter
LSU	Large subunit
LB	Luria broth
ug	Microgram

µl	Microliter
mAmp	Milliampere
min	Minute
mg	Milligram
ml	Milliliter
mm	Millimeter
mM	Millimolar
M	Molar
ng	Nanogram
nm	Nanometer
OD	Optical density
PAA	Polyacrylamide
pmole	Picomole
PS I	Photosystem I
PS II	Photosystem II
RC	Reaction center
RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	Rotations per minute
rRNA	Ribosomal nucleic acid
RT	Room temperature (18°C)
s	Second
SDS	Sodiumdodecyl sulfate
tRNA	Transfer ribonucleic acid
v	Volume
V	Volts
w	Weight

TABLE OF CONTENTS

		PAGE
1.0	Introduction	1
2.0	Materials and methods	12
2.1	Chemicals	12
2.2	Buffers	13
2.3	Sterile conditions	15
2.4	Siliconization	15
2.5	Bacterial growth and media	16
2.6	Molecular cloning	17
2.6.1	Cloning of the 32 kd gene from Amaranthus retroflexus	18
2.6.2	Recombinant plasmids	20
2.7	Growth and harvesting of bacterial cells	21
2.7.1	Large-scale plasmid DNA isolation by lysis with Triton X-100	22
2.7.2	CsCl/EtBr gradients	22
2.7.3	Large-scale isolation of plasmid DNA by alkali lysis	23
2.8	Small scale isolation of DNA	24
2.9	Restriction enzyme digestion	25
2.10	Gel electrophoresis	26
2.10.1	Agarose gel electrophoresis	26
2.10.2	Polyacrylamide gels	26
2.10.2.1	Non-denaturing polyacrylamide gels	27
2.10.2.2	Denaturing polyacrylamide gels	28
2.10.3	RNA electrophoresis in formaldehyde/agarose gels	28
2.11	DNA fragment preparation	29
2.11.1	Electroelution of DNA from agarose or PAA gels	29
2.11.2	Salt extraction of DNA from PAA gels	30
2.11.3	DNA purification by column chromatography	30
2.12	Hybridization techniques	31
2.12.1	Southern and Northern transfers	31
2.12.2	Radiolabeling of DNA	32
2.12.3	DNA hybridization to nitrocellulose-bound nucleic acids	34

2.12.4	Grunstein-Hogness colony screening	35
2.13.1	DNA sequencing by base-specific chemical clevage	36
2.13.2	Computer analysis of DNA sequences	38
2.13.3	RNA secondary structure	40
2.14	S1 nuclease analysis	40
3.0	Results	42
3.1	Location of genes for polypeptides of the reaction center core of the PS II complex	42
3.2	Sequencing strategy	42
3.3	P680 chlorophyll *a* apoprotein gene-restriction map	42
3.4	Fragment preparation for sequencing	44
3.5	DNA sequence of the P680 chlorophyll *a* apoprotein gene	48
3.6	Amino acid sequence of the P680 chlorophyll *a* apoprotein	62
3.7	Transcription of the P680 chlorophyll *a* apoprotein gene	64
4.0	Location and sequencing stradegy for the 44 kd chlorophyll *a* apoprotein gene from PS II	69
4.1	Nucleotide sequence for the 44 kd chlorophyll *a* apoprotein gene	71
4.2	Amino acid sequence of the 44 kd chlorophyll *a* apoprotein	80
5.0	The 32 kd herbicide-binding protein gene from *Amaranthus*	83
5.1	Cloning of the 32 kd gene from a herbicide-resistant strain of *Amaranthus retroflexus*	85
5.2	Restriction map of pMHamE	85
5.3	DNA sequencing of the 32 kd gene	85
5.4	Amino acid sequence of the 32 kd gene from herbicide-resistant *Amaranthus retroflexus*	93
5.5	The tRNA HIS gene	93

6.0	Discussion	97
7.0	Summary	113
8.0	Literature	115

ACKNOWLEDGMMENTS

I would like to especially thank Prof. Dr. R.G. Herrmann for his support and guidance during this project as well as the opportunity to carry on many exciting scientific discussions. I would also like to extend my special gratitude to Dr. Achim Trebst (Ruhr Universität, Bochum) and to Dr. Wilfried Draber of Bayer AG for their enthusiasm and financial assistance; in particular, for their patience, friendliness, hard work, and helpful suggestions I am grateful to Barbara Schiller, Inge Renner, Thomas Rothe and the other members of the Botanik IV Institut.

I am thankful for the financial support from the Graduiertenförderung, Prof. Herrmann (DFG), and Bayer AG in Leverkusen.

1.0 Introduction

Photosynthetic organisms convert about 10^{11} tons of carbon into organic compounds every year. One-third of this carbon fixation is carried out by microorganisms in the oceans. The processes of photosynthesis can be subdivided into light and dark phases. In the light reactions the quantized energy of sunlight is absorbed by "antenna" molecules and transferred by pi electron resonance to a reaction center chlorophyll molecule. In its excited state, the reaction center chlorophyll is a very strong reductant and promptly releases an electron. Oxidized chlorophyll is reduced ultimately by extraction of an electron from a third molecule which is water in higher plants. The reduction of carbon dioxide to carbohydrates occurs in a reaction not requiring light.

In higher and lower plants, the light reactions of photosynthesis take place in a specialized biomembrane, the thylakoid membrane, which is located in the chloroplast. The enzymes responsible for the actual fixation of carbon dioxide and synthesis of carbohydrates are soluble proteins located in the organelle's stroma.

The largest group of prokaryotic photosynthetic organisms, the blue-green algae (cyanobacteria), perform an oxygenic photosynthesis which is mechanistically identical with the chloroplast-mediated process in green plants. Photosynthetic bacteria do not evolve oxygen, but the photochemical reactions they use for the capture of light energy are similar to those which occur in green plants.

The chlorophyll or bacteriochlorophyll which is oxidized to produce a cationic free radical is bound to proteins in complexes called the reaction center. The removal of the electron from the aromatic pi-electron system causes the loss of strong absorption bands at long wavelengths. Measurements of changes in the optical absorption spectrum were first made in the 1950's by L. Duysens (1952) who found that illuminated purple photosynthetic bacteria show a small absorbance decrease at 870 nm. The bacteriochlorophyll reaction center complex

was accordingly named P870. In chloroplasts, similar observations by Kok (Kok et al., 1961) lead to the discovery of photosystem I with the reactive chlorophyll complex P700, and P680, the reaction center of photosystem II was subsequently discovered by H. Witt and colleagues. In recent years, reaction centers have been purified from both photosynthetic bacteria and plants by conventional detergent protein purification techniques. Reaction centers isolated from the purple nonsulfur bacterium Rhodopseudomonas sphaeroides contain three hydrophobic subunits L,M, and H, in a 1:1:1 ratio (Okamura et al., 1974; Steiner et al., 1974). The cofactors are four bacteriochlorophylls, two bacteriopheophytins, two ubiquinones and one iron molecule. The amino-terminal amino acids of the three subunits from R. sphaeroides have been determined by Edman degradation (Sutton et al., 1982). The DNA sequence and the deduced amino-terminal amino acids for R. capsulata subunits show 75% conservation between the two species (Youvan et al., 1984).

Native photoreceptor units from R. viridis, obtained by sucrose density gradient centrifugation contain a central core thought to contain the reaction center complex, surrounded by a ring of protein which probably serves a light harvesting function (Jay et al., 1984). The reaction center contains four polypeptides designated cytochrome, H, M, and L and the protein ring is composed of three subunits.

The primary structure of the M subunit of the reaction center from R. sphaeroides was recently deduced from the DNA sequence of the corresponding gene (Williams et al., 1984). The M subunit is composed of 307 amino acids (M_r=34,265). Labeling techniques show the M subunit spans the cytoplasmic membrane (Bachmann et al., 1981; Valkirs and Feher, 1982). The amino-acid sequence deduced from the gene was examined for membrane-spanning regions by the method of Kyte and Doolittle (1982). Five hydropathic segments longer than twenty residues are present. Circular dichroism and polarized infrared spectroscopy data also indicate that the reaction center protein is largely in an alpha helix configuration perpendicular to the plane of the membrane (Nabedryk et al., 1982). Since the number of spanning segments is not even, the amino and carobxyl ends of the protein must be on opposite sides of the membrane.

Bacterial reaction centers contain two bacteriochlorophylls and two bacteriopheophytins in addition to the pair of bacteriochlorophylls that form P870. The initial charge-transfer state of the reaction center decays by transfer of an electron from the bacteriochlorophyll to one of the bacteriopheophytins. Further electron transfer involves two quinones. While these reactions are occurring on the reducing side of the reaction center, a c-type cytochrome replaces the electron that was removed from P870 priming the reaction center for the next excitation. Upon a second excitation, a second electron is transferred through P870, bacteriochlorophyll to bacteriopheophytin and the first quinone to the second quinone. The second quinone is now fully reduced and is transformed to the dihydroquinone by the uptake of two protons from the cytoplasm. These protons are released outside the cell when the dihydroquinone is reoxidized.

Bacterial electron transfer is cyclic (Figure 1). The electrons transferred to the second quinone return to P870 through a chain of additional electron carriers including a b-type cytochrome, another molecule of ubiquinone, an iron-sulfur protein and a c-type cytochrome communicating with P870. The cyclic flow of electrons is responsible for the generation of a transmembrane electrochemical potential gradient such that the inside of the cell is negatively charged. The movement of protons into the cell is linked to the production of ATP and regulated by an ATPase similar to that of the mitochondrial inner membrane.

While quite a bit is known about bacterial photosynthetic electron transport, much less is known about the considerably more complex electron transport occuring in higher plants. Higher plants contain two types of reaction centers, PS I and PS II, containing the chlorophylls P700 and P680, respectively. Electron flow is mostly linear but can be cyclic around PS I. The oxidized components are reduced by extracting electrons from water. PS II and PS I are connected in series as shown in the Z scheme first suggested by Hill and Bendall (1960). Figure 1 also summarizes the linear photosynthetic electron flow occurring in higher plants.

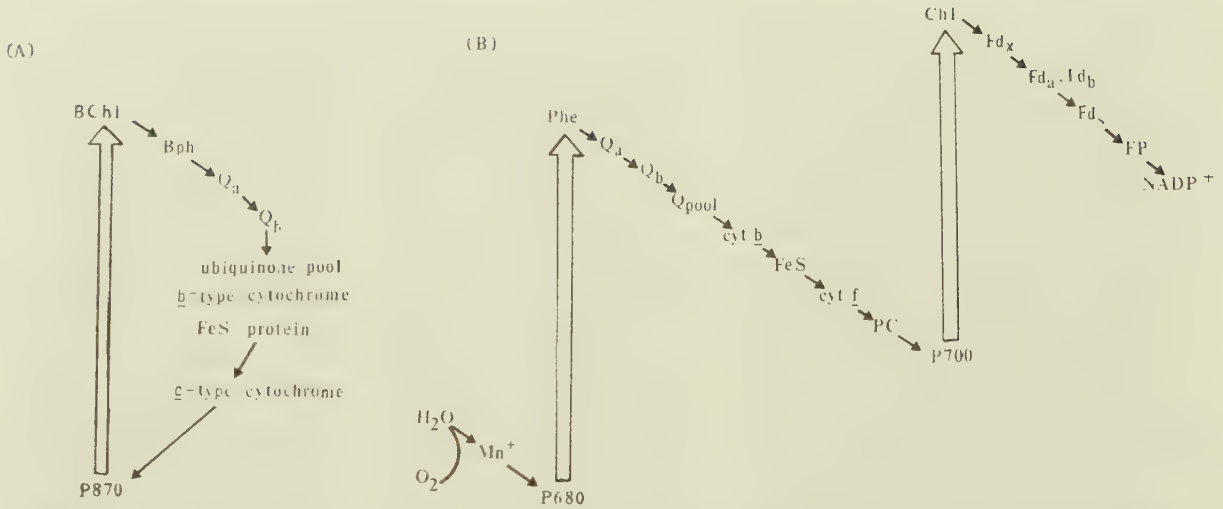


Figure 1: (A) Scheme of cyclic electron flow in purple bacteria. (B) Z scheme of linear electron transport in thylakoid membranes of the chloroplast organelle. Electron carriers are positioned vertically according to their E_m values with the most negative E_m values at the top. Downward flow is thermodynamically spontaneous. Charge separation within the reaction centers P870, P680 and P700 is indicated by thickened lines. BChl= bacteriochlorophyll a, Bph= bacteriopheophytin a, Phe= pheophytin, Q_a and Q_b= ubiquinones in (A) and plastoquinones in (B), Q_{pool}= plastoquinone pool, cyt b= cytochrome b_6, FeS= Rieske iron-sulfur protein, cyt f= cytochrome f, PC=plastocyanin, Chl= chlorophyll a, Fd_x, Fd_a, and Fd_b= ferrodoxin proteins, Fd_s= soluble ferrodoxin, FP= flavoprotein, NADP= nicotinamide adenine dinucleotide phosphate.

Six supramolecular complexes containing at least forty proteins are involved in the primary and associated electron transport processes of green plant photosynthesis localized in the thylakoid membranes, four of which are associated with chlorophyll.

(1) The PS I complex isolated from spinach (Bengis and Nelson, 1977; Nelson and Notsani, 1977; Westhoff et al., 1983a) contains eight polypeptides: two P700 apoproteins of similar molecular weight (67 kd) and smaller polypeptides of 22, 18.5, 18, 16, 12 and 10 kd. About forty chlorophyll a molecules and one ß carotene molecule are associated with the complex which is located in the stromal lamallae.

(2) The PS I associated chlorophyll a/b light harvesting complex contains two chlorophyll a binding proteins which bind about 30% of the total chlorophyll a (Mullet et al., 1982), carotenoids and all of the chlorophyll b.

(3) The PS II complex isolated from various higher plants contains at least eight polypeptides of molecular weights 51, 44, 34, 32, 32, 23, 16, and 11 kd (Mullet and Arntzen, 1981; Akerlund et al., 1983; Metz and Miles, 1982; Yamamoto et al., 1983). Chlorophyll a and ß carotene are associated with the two largest subunits (Delepelaire and Chua, 1981). Recent evidence suggests that the 51 kd protein bears the reaction center pigment P680 involved in PS II primary charge separation (Camm and Green, 1983; Nakatani et al., 1984). The herbicide-binding 32 kd protein located at the reducing side of the complex probably contains the plastoquinone binding site (Steinback et al., 1981). A similar protein, D2, has been found in Chlamydomonas (Rochaix et al., 1984). Cytochrome b_{559} probably contains an iron atom involved in the transfer of electrons between the two quinones. While the 51, 44, two 32 kd proteins and cytochrome b_{559} are integral membrane proteins located in the grana, the loosely associated 34, 23, and 16 kd components are located in the thylakoid lumen and probably participate in water photolysis.

(4) PS II chlorophyll a/b light harvesting complex contains two to five related subunits of 25 kd. They bind 50% of the total chlorophyll a

and almost all of the chlorophyll b (Anderson et al., 1981).

(5) The cytochrome b_6/f complex contains at least four subunits (Hurt and Hauska, 1981): cytochrome f (33 kd), cytochrome b_6 (23 kd), Rieske iron-sulfur protein (19 kd) and subunit 4 (15.2 kd). Both cytochromes are probably present in only one copy per complex. Cytochrome b binds two heme molecules (Widger et al., 1984) while the high-potential Rieske protein carries two non-heme iron molecules (Hurt and Hauska, 1981, 1982). The complex also contains bound plastoquinol, and when incorporated into phospholipid vesicles can function as an electrogenic pump (Hurt and Hauska, 1981, 1982; Hurt et al., 1982). It is distributed in both granal and intergranal regions of the thylakoid membrane (Cox and Andersson, 1981).

(6) The ATP synthase complex, like its bacterial and mitochondrial counterparts, contains eight subunits. The hydrophilic catalytic part CF_1 which sits on top of the thylakoid membrane, contains five subunits designated alpha, beta, gamma, delta, and epsilon. The hydrophobic CF_0 part which builds a channel for proton flow through the membrane has three subunits designated I, II, and III. The now widely accepted subunit stoichiometry for the thylakoid ATP synthase is 2 alpha, 2 beta, 1 gamma, 1 delta, 2 epsilon, 2 CF_0-I, 2 CF_0-II, and 6 CF_0-III (Nelson, 1981). The six subunit III molecules probably form the channel for proton flux (Nelson, 1982).

The components of the thylakoid membrane complexes are depicted schematically in Figure 2.

The initial electron acceptors on the reducing side of P680 are similar to those of the purple photosynthetic bacterial reaction center: pheophytin a followed by two plastoquinone molecules. The primary photochemical charge separation in PS II consists in the transfer of an electron form the primary electron donor P680 chlorophyll a to a pheophytin a molecule (Klimov et al., 1977). The P680 apoprotein has recently been identified as the 51 kd species of the PS II particle (Camm and Green, 1984; Nakatani et al., 1984). The secondary electron

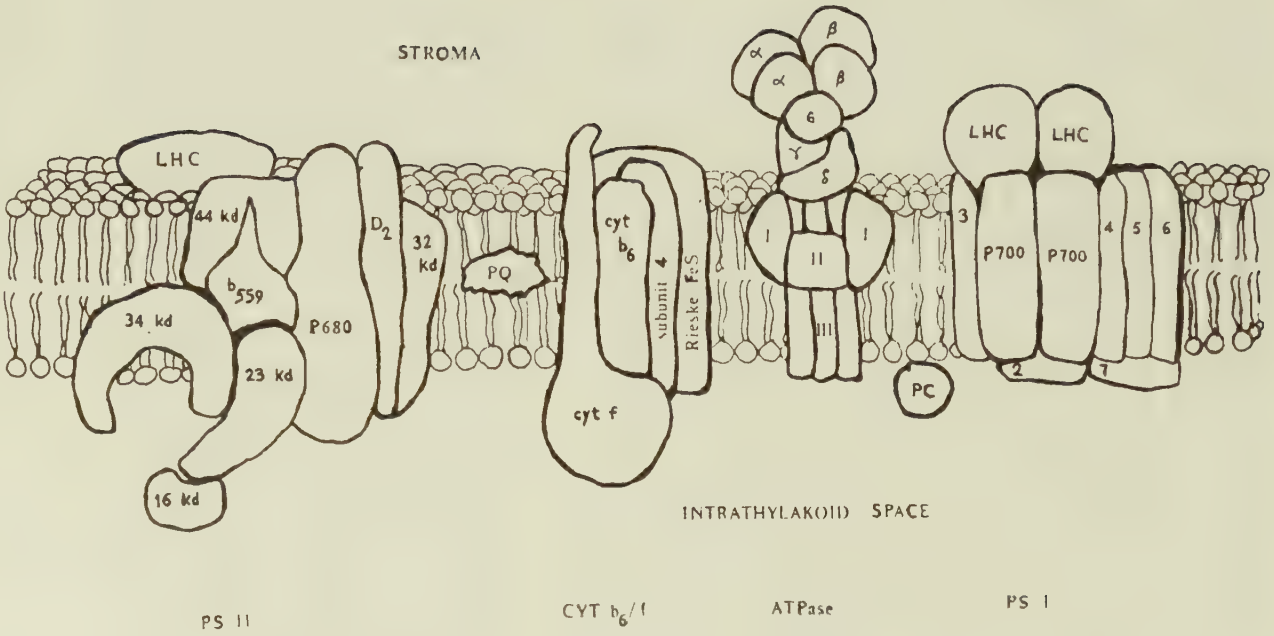


Figure 2: Cartoon showing the components of the supramolecular complexes in the thylakoid membrane of the chloroplast organelle.

acceptor is a tightly bound plastoquinone molecule whose reduction results in the transfer of an electron to a second more loosely bound plastoquinone molecule. As in bacteria, the second plastoquinone goes through a two-step reduction, the first electron transfer forming a relatively stable semiquinone anion and the second electron transfer forming the fully reduced quinone. After protenation, the quinone is released from its protein binding site into the lipid soluble phase of the membrane. Freely diffusing plastoquinone molecules comprise a pool and these molecules act as a shuttle to transfer electrons to the cytochromes and photosystem I. In chloroplasts, electron transport is coupled via the $\underline{b_6/f}$ complex to proton movement across the thylakoid membrane into the intrathylakoid space. The resulting proton-motive force drives the synthesis of ATP by the ATP synthase catalytic unit which is exposed to the stromal side of the thylakoid membrane. At the same time in separate reactions, water is oxidized, $NADP^+$ reduced, and oxygen liberated. Depending on the stromal ratio of the concentration of NADPH to $NADP^+$ electrons will flow in cyclic photophosphorylation between the cytochrome complex and photosystem I. Considerably more ATP is produced by cyclic flow than by passage of electrons from water to $NADP^+$. The relative activities of cyclic and linear electron transport flows determine how much light energy is converted into ATP and NADPH, respectively.

The biogenesis of the chloroplast including the formation of the photosynthetic thylakoid membrane involves a cooperative interaction between plastid and nuclear genomes and requires protein synthesis in both plastids and cytosol (Herrmann et al., 1985). The genetic information in plastids is contained on a circular chromosome with an average size of between 100 and 200 kb depending on the plant species. Physical maps of restriction endonuclease sites in plastid DNA have been constructed and in many cases individual restriction fragments cloned into appropriate vectors. These analyses have shown that the majority of plastid DNAs contain an invert repeat sequence separated by a large and a small single copy region. Plastid chromosomes could code for at least 150 proteins of 20,000 kd size. They contain coding sequences for components of the organelle translational machinery including 23S, 16S

and 5S rRNA and, about forty species of tRNA and several ribosomal proteins (e.g.Herrmann and Possingham, 1980; Whitfeld and Bottomley, 1983;Bohnert et al., 1982).

In spinach, the location of genes for thylakoid membrane proteins has been established using two complementary techniques, hybrid-selection translation and DNA-programmed coupled transcription-translation in an S30 cell-free lysate from E. coli (Zubay et al., 1970). In both instances the products of translation were identified immunologically. Using these techniques, the location of the genes for four ATP synthase subunits alpha, beta, epsilon and subunit III of the CF_0 (Westhoff et al., 1981; Alt et al., 1983), for the cytochromes b_6 and f, and subunit 4 of the cytochrome complex (Alt et al., 1983), for the P680 apoprotein (51 kd), the 44 kd apoprotein, D2, the cytochrome b_{559} of the PS II reaction center (Westhoff et al., 1983; Alt et al.,1984), and the two P700 chlorophyll a apoproteins of the PS I reaction center (Westhoff et al., 1983; Kirsch et al., 1985) have been determined as well as the herbicide-binding 32 kd polypeptide associated with PS II (Driesel et al., 1980) and the large subunit of ribulose bisphosphate carboxylase/oxygenase (Herrmann and Possingham, 1980; Whitfeld and Bottomley, 1980) on the spinach plastid chromosome. The location of genes for thylakoid membrane polypeptides on the plastid chromosome from spinach is summarized in Figure 3 (Westhoff et al., 1983).

Nuclear coded proteins are synthesized in the cytoplasm then imported through the chloroplast membrane into the stroma (Parthier, 1982). These include the delta and gamma subunits of CF_1, subunit II of the CF_0 of the ATP synthase, the Rieske FeS protein of the cytochrome complex as well as subunits II and III of the PS I complex (Herrmann et al., 1985), constituent polypeptides of the chlorophyll a/b light harvesting complexes of both photosystems and three loosely associated polypeptides of PS II (34, 23, and 16 kd), probably involved in the splitting of water (Westhoff et al., 1984). Plastocyanin, ferredoxin and ferredoxin reductase appear to be coded for in the nucleus (Westhoff et al., 1981; Alt et al., 1983; Alt et al., 1984; Westhoff et al., 1983; Schmidt et al., 1981; Mullet et al., 1982).

The fact that the plastid chromosome represents a major genetic determinant for proteins of the thylakoid membrane and that the genes are accessible to analysis using DNA sequencing and cloning techniques has made it possible to investigate the molecular basis for the control of expression of organellar genes and to gain a deeper understanding of the mechanisms responsible for the coordinate expression of nuclear and organellar genes during various stages of differentiation. The fine structure analysis of genes for reaction center polypeptides also permits the dissection of the process of photosynthetic energy transduction and electron transport at the molecular level.

The purpose of this study was the analysis of genes for polypeptides of the watersplitting PS II complex, specifically those for the chlorophyll conjugated apoproteins involved in the primary photosynthetic reactions and the herbcide binding 32 kd protein located at the reducing side of the complex. This latter protein is of particular interest since a large group of artificial chemical inhibitors of photosynthetic electron transport, including the triazine herbicides, act at the same site on the photosynthetic membrane; electron transport is blocked at a protein-plastoquinone complex which functions as the second stable electron acceptor of PS II. The high-affinity binding site has been localized to membrane polypeptides of 34 and 32 kd by photoaffinity labeling (Pfister et al., 1981; Gardner, 1981). Chloroplasts of triazine-susceptible and triazine resistant biotypes of Amaranthus hybridus synthesize the 32 kd polypeptide but the protein from the resistant biotype does not bind the herbicide (Steinback et al., 1981). Sequencing of the gene for this polypeptide from the two biotypes of Amaranthus revealed that they differ by a single amino acid change (Hirshberg and McIntosh, 1983). In this study a 32 kd protein gene from a herbicide resistant variety of Amaranthus retroflexus was cloned and sequenced in order to find the point mutation responsible for the herbicide resistance.

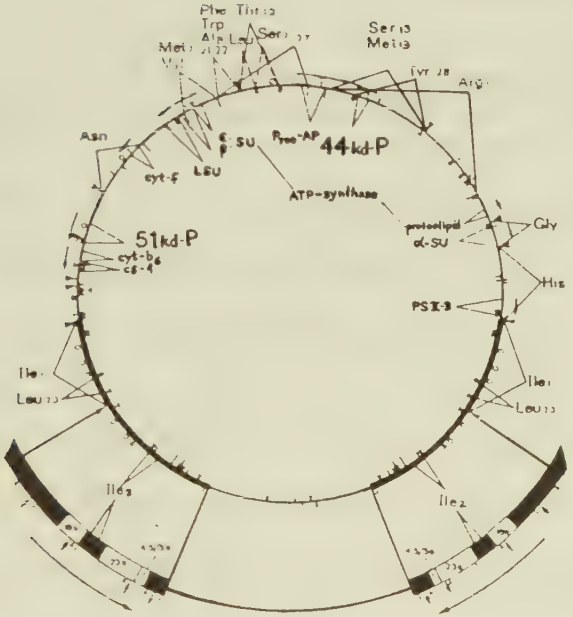


Figure 3: Physical map showing the location of the genes for the P680 chlorophyll *a* apoprotein (51 kd-P), the 44 kd chlorophyll *a* apoprotein (44 kd-P) and the 32 kd protein (PS II-3) associated with the PS II reaction center core, on the spinach plastid chromosome (Westhoff et al., 1983; Driesel et al., 1980). Genes for the two P700 apoproteins of PS I (P700-AP), cytochromes *f* and b_6, subunit 4 (CS-4), ATP synthase subunits alpha, beta, epsilon and CF_0-III (α, β, ϵ-SU and proteolipid) and the large subunit of ribulose bisphosphate carboxylase oxygenase (LSU) are also indicated. The inverted repeat with rRNA genes is indicated by thicker lines and tRNA genes by the name of their associated amino acid. Transcription polarities are indicated by parallel arrows. Restriction site symbols follow Westhoff et al., 1983.

2.0 Materials and methods

2.1 Chemicals

Acrylamide, Methylenebisacrylamide, Amberlite MB 1, Ammoniumpersulfate, Ethidium bromide, Triton X-100, Polyethyleneglycol 6000, Xylenecyanol, Calfs thymus DNA (Serva Feinbiochemica, Heidelberg)

Seakem ME agarose (FMC Corp. Rockland, USA)

Piperidine, Sodium acetate, 2-Butanol, Boric acid (Ferak, Berlin)

Dimethylsilane, Phenol, Chloroform (Fluka AG, Chem. Fabrik, Buchs, Switzerland)

Sephacryl S-200 Superfine (Pharmacia Fine Chemicals, Uppsala, Sweden)

Dimethylsulfate, Cesium chloride (Janssen Chimica, Beerse, Belgium)

Dithiothreitol, Ampicillin, Bovine serum albumin (Sigma Chemical Company, St. Louis, USA)

Mercaptoethanol, Urea (Carl Roth, Karlsruhe)

Hydrazine (Eastman Kodak Co., Rochester, N.Y. USA)

Bactotryptone, Yeast extract, Case amino acids (Difco Laboratories, Detroit, Michigan, USA)

Radioisotopes: (α-^{32}P)-dATP and (α-^{32}P)-dCTP, 3000 Ci/mmol, 10 mCi/ml (Amersham Buchler, Braunschweig; NEN Nuclear) (γ-^{32}P) ATP, 3000 Ci/mmol, 10 mCi/ml (Amersham Buchler, Braunschweig)

tRNA (*E. coli*), Deoxyribonucleotides, Hind III digest of lambda DNA, RNase, Lysozyme, DNA-polymerase large fragment (Klenow fragment), E. coli DNA polymerase I, Deoxyribonuclease I, Calves intestinal phosphatase, T4 polynucleotide kinase, T4 polynucleotide ligase,

S1 nuclease, Proteinase K, Eco RI, Hind III, and Bam HI (Boehringer Mannheim, Germany)

Urea, Bam HI, Eco RI, Hae III, Hha I, Hinc II, Hinf I, Taq I (BRL , Neu-Isenburg)

Hae III digest of ØX-174 DNA, Alu I, Ava I, Bam HI, Dde I, Hae III, Hha I, Hinc II, Hind III, Hinf I, Hpa II, Pvu I, Rsa I, Sau 3A, Sal I, Taq I, Xba I (New England Biolabs, Schwalbach, TS)

Other chemicals of analytical quality (Merck, Darmstadt)

2.2 Buffers

TE

10 mM	Tris-HCl, pH 8.0
1 mM	Na_2-EDTA

TBE

100 mM	Tris-borate, pH 8.3
2.5 mM	Na_2-EDTA

TAP

40 mM	Tris-acetate, pH 7.5
20 mM	Na-acetate
5 mM	Na_2-EDTA

Salt extraction buffer

0.5 M	NaCl
10 mM	$MgCl_2$
10 mM	Tris-HCl, pH 8.0
1 mM	Na_2-EDTA

Agarose sample buffer

50% v/v	Glycerol
0.1% w/v	Bromophenol blue

DNA sequencing sample buffer

50 mM	Tris-borate, pH 8.3
1.25 mM	Na_2-EDTA
80% v/v	Formamide (deionized)
0.1% w/v	Bromophenol blue
0.1% w/v	Xylene cyanol
1 mM	Na_2-EDTA

PAA sample buffer

50% v/v	Glycerol
0.1% w/v	Bromophenol blue
0.1% w/v	Xylene cyanol

SSPE

0.18 M	NaCl
1 mM	EDTA
10 mM	NaH_2PO_4, pH 7.4

STE

0.1 M	NaCl
10 mM	Tris-HCl, pH 7.8
1 mM	EDTA

MOPS

20 mM	Morpholinosulfonate, pH 7.0
5 mM	Na-acetate
0.1 mM	EDTA

SSC

0.15 M	NaCl
0.015 M	Na-citrate

CTA Stop

0.3 M	Na-acetate
0.1 mM	Na_2-EDTA
25 ug/ml	tRNA

DMS Buffer

50 mM	Na-cacodylate
10 mM	$MgCl_2$
0.1 mM	Na_2-EDTA

G Stop

1.5 M	Na-acetate
1.0 M	ß-mercaptoethanol
100 ug/ml	tRNA

S1 nuclease buffer

50 mM	Na-acetate, pH 4.6
1 mM	$ZnSO_4$
0.25 M	NaCl
20 ug/ml	Calfs thymus DNA (sheared)

Nick translation buffer

50 mM	Tris-HCl, pH 7.2
10 mM	$MgSO_4$
0.1 mM	DTT

2.3 Sterile conditions

Glassware, solutions and transfer vehicles were sterilized by autoclaving at 120°C under 1.2 Atm pressure for 30 to 45 minutes or by application of dry heat (150 to 200°C) overnight. All bacterial work was carried out under a laminar flow hood.

2.4 Siliconization

To prevent adsorption of DNA to glassware and facilitate removal of gels from their molds, glass surfaces were treated with a solution of 1% dimethylsilane in carbon tetrachloride, and vapors evaporated under a

fume hood to remove any acidic residue.

2.5 Bacterial growth and media

The C600 strain used in these studies is an E. coli derivative of K12 sensitive to bacteriophage lambda (Appleyard, 1954). The stock culture and recombinant derivatives were stored in a 50% glycerol containing buffer (Storage medium) at -20°C. Strains were usually grown in Luria broth and for large scale plasmid isolations in case amino acids media. Media were supplemented with antibiotics as indicated. Cell density was determined by measuring the scattering by the cells of the collimated beam from a spectrophotometer at a wavelength of 550 nm. Under these conditions 1 OD unit corresponds to approximately 8 x 10^8 cells/ml.

(1) Bacterial storage medium

3 gm	KH_2PO_4
7 gm	K_2HPO_4 $2 \cdot H_2O$
0.5 gm	Na_3-citrate $2 \cdot H_2O$
0.1 gm	$MgSO_4$
500 ml	Distilled water
500 ml	Glycerin

(2) Luria broth (LB)

10 gm	Bactotryptone
5 gm	Yeast extract
5 gm	NaCl
3 dps	32% NaOH /per liter

(3) Case amino acids medium

5 gm	Case amino acids
2 mg	Thiamin-HCl
6 gm	NaH_2PO_4
3 gm	KH_2PO_4
0.5 gm	NaCl
1 gm	NH_4Cl
1 ml	$MgSO_4$
0.72 %	Glucose
1 ml	0.1 M $CaCl_2$
2 gm	Uridine /per liter

For agar plates add 16 gm Bacto-agar per liter.

(4) Antibiotics

Antibiotics were prepared as stock solutions, sterile filtered and stored at 4°C. After autoclaving media were allowed to cool to 45°C before adding the antibiotics and pouring plates.

ANTIBIOTIC	MEDIA/PLATE CONCENTRATION	PLATE USAGE PERIOD
Ampicillin	40 ug/ml	1-2 weeks
Tetracycline*	15 ug/ml	1 month
Chloramphenicol	10 ug/ml	5 days

* Tetracycline stock was made up in 50% ethanol and stored protected from light.

2.6 Molecular cloning

For molecular cloning, the foreign DNA containing the relevant gene(s) is cleaved with a restriction endonuclease which creates single-stranded or blunt ends. A vector containing a similar or compatible cleavage site

in a unique and selectively marked site, such as an antibiotic resistance gene, is similarly cleaved. The two DNAs are covalently joined by a ligase enzyme and the recombinant plasmid used to transform bacteria. Simple recircularization of the vector DNA can be limited to some extent by adjusting the concentrations of foreign DNA and vector DNA during the ligation reaction. However, in shotgun cloning experiments the treatment of the vector with calves intestinal phosphatase which removes the 5' phosphates was found to be easier and more effective in preventing recircularization. The dephosphorylated vector DNA and the undephosphorylated plastid DNA were joined using T4 ligase and the resulting mixture of linear and circular molecules used to transform competent bacteria. Transformed bacteria were detected on selective antibiotic media and screened for the presence of passenger DNA by replica plating. Those clones containing foreign DNA were further screened by hybridization to an appropriate probe to detect bacteria harboring the particular gene of interest.

2.6.1 Cloning of the 32 kd gene from Amaranthus retroflexus

(1) Preparation of plastid DNA fragments

Plastid DNA was kindly isolated according to Herrmann (1982) by Ms. Barbara Schiller. The gene was located on a 9.6 kb Sal I fragment from the plastid DNA (Results). Three hundred micrograms of plastid DNA from the herbicide resistant strain of Amaranthus was digested with Sal I, the fragments separated on 0.8% agarose gel and those between 7 and 10 kb isolated by electroelution. This mixture was then digested with Eco RI giving between 30 and 40 fragments in preparation for ligation.

(2) Preparation of pBR325 vector DNA

Five micrograms of pBR325 DNA which had been cleaned over CsCl/EtBr gradients was digested with Eco RI, extracted with phenol-chloroform (1:1), chloroform alone, then ether. The DNA was precipitated, resuspended in 50 µl of 50 mM Tris-HCl, pH 9.0, 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$, 1 mM spermidine. 0.1 units of calves intestinal phosphatase, enough to remove the terminal phosphates from 10 pmole of 5' ends was added; the 5 µg of linear pBR325 (6 kb length) has 2 pmole of 5' ends. After a 30 minute incubation at 37°C, 40 µl of water was added, 10 µl of 100 mM Tris-HCl,

pH 8.0, 1M NaCl, 10 mM EDTA, and 5 μl of 10% SDS. This mixture was heated to 68°C for 15 minutes to inactivate the phosphatase, then the enzyme removed by three successive extractions with chloroform-phenol (1:1). The DNA was precipitated with ethanol, recovered by centrifugation, washed with 70% ethanol, dried under vacuum and resuspended in 5 μl of sterile distilled water.

(3) Ligation of Eco RI fragments and pBR325

The ligation was performed in a total of 10 μl containing 0.66 M Tris-HCl, pH 7.6, 50 mM $MgCl_2$, 50 mM DTT, 10 mM ATP, 1 unit of T4 DNA ligase, 1 μg of phosphatase-treated Eco RI cut pBR325 and 500 to 1000 ng of Eco RI fragments from herbicide-resistant plastid DNA. The end concentration of DNA was 1.5 to 2.0 μg/μl. Ligation was allowed to proceed overnight (16 hours) at 12°C then the ligation mixture was diluted to a DNA concentration of 15 to 20 ng/μl with TE buffer.

(4) Preparation of transformation competent bacteria

When the optical density of a 30 ml culture of C600 bacteria reached 0.5 (early log phase) the culture was transferred to a sterile pre-cooled 50 ml tube and the cells gently pelleted for 5 minutes (4080 g, HB 4 rotor) at 4°C. The white pellet of 1 cm diameter was gently resuspended in 20 ml of ice-cold 0.1 M $MgCl_2$, left on ice for 20 min, pelleted as before and resuspended in 20 ml of 0.1 M $CaCl_2$. After 20 min on ice the cells were collected as before, resuspended in 1.0 ml of 0.1 M $CaCl_2$ and left at 4°C overnight. During this time, the competence rises by a factor of six (Dagert and Ehrlich, 1979).

(5) Transformation

The transformation was performed by mixing 200 μl of competent bacterial cells and 3 μl of ligation mixture (45 to 60 ng DNA), allowing this mixture to sit on ice for 30 min then transferring it to a 42°C water bath ("heat shock") for 2 min before returning it to ice. One ml of LB was added to start growth and the cells incubated for 1 hour at 37°C with 170 rpm rotary shaking. The cells were then plated in 150 μl aliquots on LB plates supplemented with 10 μg/ml ampicillin to select for transformants.

(6) Grunstein-Hogness screening of transformants

Transformants containing an insert, as identified by their inability to grow on LB plates supplemented with 10 μg/ml chloramphenicol, were screened according to the Grunstein-Hogness procedure (Section 2.12.4) using as hybridization partner a ^{32}P-labeled Pst I- Hind III fragment (250 bp) containing sequences from the spinach 32 kd gene (Zurawski et al.,1982).

(7) Purification of recombinant clones

A sterile toothpick was used to transfer the desired recombinant bacteria to 3 ml of LB containing 10 μg/ml ampicillin and the cells serially diluted to a concentration of 10,000 per ml. Twenty microliters of the diluted bacteria were spread on LB plates containing ampicillin and the bacteria allowed to grow overnight at 37°C. A single colony was used to inoculate 30 ml of LB supplemented with 10 μg/ml ampicillin, the culture grown overnight at 37°C, with 170 rpm shaking and the cells then aliquoted into 1 ml portions in sterile Eppendorf tubes. After a 3 min centrifugation, the clear supernatant was removed and the pellet resuspended in Storage medium and stored at -20°C or -70°C.

2.6.2 Recombinant plasmids

(1) pWHsp 207 and pWHsp 208

The non-conjugative pBR322 (Bolivar et al., 1977) carrying resistance genes for ampicillin and tetracycline within its 4361 bp circular DNA served as vector for cloning of Sal I primary fragments from spinach plastid DNA. The recominbant plasmids pWHsp 207 and pWHsp 208 carrying the 5.6 and 5.2 kbp Sal I fragments 7 and 8 , respectively, were selected from a shot-gun library (Herrmann et al., 1983; Alt et al., 1983; Winter, 1982). Fragment 7 contains the gene for the P680 chlorophyll *a* reaction center protein while the adjacent fragment 8 carries the genes for cytochrome b_6 and subunit 4.

(2) pWHsp 206/ES

A 2050 bp Eco RI-Sal I fragment excised from pWHsp 206 was ligated into the Eco RI-Sal I digested pBR322 to create the recombinant plasmid

pWHsp 206/ES (Schmidt, unpublished results). The clone contains the entire 32 kd protein gene of spinach.

(3) pWHsp 510

The vector pBR322 was opened at its single Pvu I site, treated with calves intestinal phosphatase, and ligated to the isolated Pvu 10 fragment (Herrmann and Whitfeld, 1982). The 1600 bp fragment contains the Sal I site separating the primary Sal I fragments 7 and 8.

(4) pMHamE

The plasmid pBR325 (Bolivar et al., 1978) containing the ampicillin resistance gene from pBR322 and the chloramphenicol resistance gene from a Cm transposon was used as vector for the cloning of the 32 kd herbicide-binding protein gene from Amaranthus retroflexus (Section 2.6.1).

(5) pWHsp 104, pWHsp 408b, and pWHsp 416

The three hybrid plasmids pWHsp 104, pWHsp 408b, and pWHsp 416 containing the primary fragments Pst I-4, Bam HI-8b and Bam HI-16, respectively (Herrmann et al., 1980; 1983) carry portions or all of the 44 kd chlorophyll a binding apoprotein of PS II from spinach (Westhoff et al., 1983).

2.7 Growth and harvesting of bacterial cells

Thirty mls of LB containing the appropriate antibiotic were inoculated with 0.3 ml of a stock culture in storage buffer. After overnight growth, the culture was diluted 1:100 into one liter of case amino acids media. When the OD 550 reached a value between 0.6 and 0.8, chloramphenicol was added to a concentration of 170 µg/ml (5.0 ml of 34 mg/ml in 50% ethanol) and incubation continued for 16 to 20 hours. The bacterial ribosomes are hampered by chloramphenicol and the cells cease to divide, but plasmid replication continues such that plasmids not under stringent control reach a copy number of several thousand per cell after 20 hours of growth (Clewell, 1972). Cells were always grown at 37°C with rotary shaking between 150 and 200 rpm depending on the size of the culture flask.

Cells were harvested by centrifugation at 4,000 g, 4°C in sterile 250 ml bottles for 10 min (GSA rotor) or in 1 liter bottles in a JS-4.2 swingout rotor at 4,200 rpm for 30 min then washed once with STE to remove culture media and other soluble components that might inhibit the plasmid isolation.

2.7.1 Large-scale plasmid DNA isolation by lysis with Triton X-100 (modified from Katz et al., 1973)

The bacterial pellet from a 1 liter culture was thoroughly resuspended in 20 mls of an ice-cold solution of 10% sucrose in 50 mM Tris-HCl, pH 8.0. 100 mg lysozyme (2.2×10^6 units) was added to effect a partial digestion of the β(1-4) glycosidic linkages between peptidoglycan units of the outer bacterial cell wall. Digestion was allowed to proceed for 10 min on ice. The addition of EDTA to 50 mM chelated cations with a +2 charge which hinders both further lysozyme digestion and DNases (Mg requirement). The plasma membrane was gently disrupted by the addition of Triton X-100 to a concentration of 0.18%. This non-ionic detergent saturates positively charged sites on proteins thereby leading to their denaturation and is therefore used to disrupt membrane structures, to remove proteins bound to nucleic acids and to inactivate DNases. The originally cloudy suspension becomes clear and semi-gelled as the bacterial lysis releases a tangle of high molecular weight chromosomal DNA and cell wall debris. Lysis was carried out at RT for 30 min followed by centrifugation at 14,500 g (6 x 100 rotor, 90 min). The pellet contains most of the chromosomal DNA while the supernatant contains plasmid DNA, RNA, carbohydrate and proteins. The latter was decanted and treated with 1.2 mg RNase (510 units) at 37° for 30 min to digest high molecular weight RNA, then with 8 mg of Proteinase K (160 units) for 30 min at 37°C to digest proteins. Nucleic acids were precipitated overnight by addition of NaCl to a concentration of 0.5 M and polyethylene glycol 6,000 to a concentration of 10% (Lis, 1980).

2.7. CsCl/EtBr gradients

Plasmid DNA gained by this technique was further purified by isopycnic centrifugation in CsCl/EtBr gradients. The CsCl concentration was chosen such that the gradient includes densities approximating those of the

DNA to be separated. Native superhelical DNA bands at a density of 1.59 gm/ml while relaxed DNA bands at 1.55 gm/ml. Proteins are found at the top and RNA at the bottom of the gradient. The width of the bands depends on the concentration of molecules of that density. The gradient formed in the vertical Ti50 rotor extends across the centrifugation tube and reorients from the bottom to the top during braking. An orderly reorientation was insured by removing the brake at 3,000 rpm. Gradients were prepared by dissolving the DNA in 24 ml of TE and adding solid CsCl to a concentration of 1 mg/ml. 1.9 ml of ethidium bromide (10 mg/ml in water) was added and the solution transferred to the vertical centrifugation tube. The final density of the solution is 1.55 gm/ml and can be checked by measuring the refractive index (η = 1.3860). The final concentration of ethidium bromide is 600 µg/ml. Tubes were filled and tared to within 50 mg with light parrafin oil then sealed by the application of heat. Centrifugation was carried out for 16 to 20 hours in the Ti50 rotor at 44,000 rpm at 18°C (Spinco). Under these conditions, the closed circular plasmid DNA bands just below the linear plasmid and chromosomal DNA in the middle of the gradient. The two bands were located with an ultraviolet hand lamp and removed separately with a syringe. EtBr which in the light forms radicals that induce nicking of DNA was promptly removed by sequential extraction with isoamyl alcohol. CsCl was removed by dialysis against 2 x 5l TE for 1-2 hrs each, and the DNA concentrated by extraction with n-butanol (Maniatis et al., 1982) before precipitation. DNA was collected by centrifugation at 16,000 g for 10 min (HB-4 rotor), washed with 70% ethanol and dried under vacuum. The purity of the preparation was checked by agarose gel electrophoresis. Plasmid DNA which proved resistant to restriction enzyme digestion was dialyzed more extensively against TE and reprecipitated.

2.7.3 Large-scale isolation of plasmid DNA by lysis with alkali (modified from Birnboim and Doly, 1979)

The bacterial pellet from a 1 liter culture was resuspended in 20 ml of 50 mM glucose, 25 mM Tris-HCl, pH 8.0, 10 mM EDTA containing 100 mg of freshly dissolved lysozyme. Digestion was for 5 to 15 min at RT until a decrease in the viscosity of the solution could be seen. 40 mls of a

freshly made solution of 0.2N NaOH, 1% SDS was added, gently mixed by swirling and the solution allowed to stand on ice for 10 min. SDS lyses the spheroplasts and denatures the proteins while NaOH creates a mild alkali environment (pH 12.5) which breaks the H-bonds in DNA of the bacterial chromosome as well as the plasmid. 15 mls of ice-cold 5M potassium acetate, pH 4.8, was added to neutralize the lysate. Under these conditions the closed circular plasmid molecules regain their native configuration but the E. coli chromosomal DNA remains in the denatured state. The high concentration of potassium acetate causes precipitation of protein-SDS complexes and high molecular weight RNA. All three major contaminants e.g. chromosomal DNA, high molecular weight RNA and protein were removed by centrifugation for 120 min at 19,745 g (GSA rotor). Plasmid DNA and low molecular weight RNA were recovered in the supernatant. Isopropanol (0.6 v) was added and the nucleic acids precipitated at -20°C for 30 min. The precipitate was recovered by centrifugation for 30 min at 10,444g (GSA rotor), washed once with 70% ethanol, and dried briefly under vacuum. The dried pellet was resuspended in 5 ml of TE and extracted twice with phenol saturated with TE and twice with chloroform. RNase (1.2 mg) was added and digestion allowed to proceed for 30 min at 37°C before the final precipitation with sodium acetate and ethanol. This technique allowed routine isolation of 5 to 10 mg of plasmid DNA from 6 liters of culture media.

2.8 Small scale isolation of DNA (Birnboim and Doly, 1979)

Three mls of LB containing the appropriate antibiotic was inoculated with a single bacterial colony and the culture grown to stationary phase. The cells from 1.5 ml of culture were sedimented in an Eppendorf (3 min, 1500 g). The supernatant was discarded and the pellet resuspended in 100 µl of a membrane lysing solution containing 50 mM glucose, 10 mM EDTA, 25 mM Tris-HCl, pH 8.0 and 4 mg/ml lysozyme (freshly dissolved). After 30 min on ice, NaOH and SDS were added to concentrations of 0.133N and 0.67% respectively, to effect complete lysis of the cells and denaturation of the DNA. After 5 min, 150 µl of 5 M K acetate (pH 4.8) was added to neutralize the lysate. After 60 min on ice, and 15 min in the Eppendorf centrifuge, the cleared lysate was recovered in about 400 µl of supernatant. This was immediately extracted with phenol:chloroform

(1:1), and the aqueous upper phase transferred to a new Eppendorf tube. Nucleic acids were precipitated by adding 1 ml ethanol, and cooling in a methanol/dry ice bath (-70°C) for 30 min. then recovered by centrifugation in the Eppendorf (15 min). The pellet was washed with 500 µl of 70% ethanol, dried briefly, resuspended in 0.1M Na acetate and the DNA reprecipitated with ethanol. After centrifugation, washing and drying, the pellet was resuspended in 36 µl water, 4 µl RNase (1 mg/ml in 0.1XSSC) to degrade the RNA (30 min, RT) before adding buffer and restriction endonuclease; digested plasmids were finally separated on 0.5 to 1% agarose gels.

2.9 Restriction enzyme digestion

The type II restriction endonucleases used frequently in recombinant DNA recognize a specific sequence of 4 to 6 bp which generally has a two-fold symmetry. The enzyme usually causes a double-stranded scission which may be blunt-ended or staggered, e.g. with a 3' or 5'-single-stranded extension. Restriction endonuclease digestion was used (1) to prepare DNA fragments and vectors for ligation and cloning, (2) to analyze clones containing foreign DNA inserts, (3) to analyze the physical location of restriction endonuclease cleavage sites relative to one another in a particular DNA fragment, (4) to prepare appropriate DNA fragments for radioactive end-labeling, (5) to separate a fragment with two radiolabeled ends into subfragments for sequencing.

For analytical purposes 1 ug of DNA was digested for 1 to 3 hrs with 4 to 8 units of enzyme at the appropriate temperature in a volume of 20 to 50 µl with 10 to 50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, and 1 mM DTT. Buffers contained either 100 mM NaCl, 50 mM NaCl or no NaCl. One-tenth of the appropriate sample buffer was added before electrophoretic separation of the DNA fragments on agarose or polyacrylamide gels.

To prepare DNA fragments for end-labeling, 200 to 500 µg of plasmid DNA was digested with the appropriate restriction endonuclease at a DNA concentration of 50 to 100 µg/ml. The amount of restriction endonuclease required to digest the DNA could be lowered by increasing the time of the digestion. Exonuclease contamination causing secondary

degradation was not a problem. The progress of a preparative scale DNA digestion could be tested by electrophoresis of an aliquot on a mini-agarase gel.

2.10 Gel electrophoresis

2.10.1 Agarose gel electrophoresis

One standard technique for separation, identification and purification of DNA fragments is electrophoresis through agarose gels. With the use of fluorescent intercalating dyes such as ethidium bromide, the progress of electrophoretic separation can be examined with an ultraviolet light source.

Agarose gel electrophoresis of DNA fragments was carried out with horizontal slab gels of a standard size 15 x 20 x 0.5 cm or a mini-gel size of 10 x 10 x 0.3 cm with appropriate sample wells. The running buffer was TAP supplemented with 0.5 μg/ml EtBr. The agarose was dissolved in buffer by bringing the gel solution to a boil, then cooled to 50°C before pouring into the plexiglas mold. Combs were positioned to allow 0.5 to 1 mm of agarose between the bottom of the teeth and the base of the gel so that each sample well was completely sealed. Gels were allowed to set at lease one hour before use. Standard size gels were run at constant voltage of 5 V/cm and mini-gels covered with running buffer at 15 to 20 V/cm. Samples were supplemented with one-tenth volume of a sample buffer containing 50% glycerol (v/v) and 0.1% bromophenol blue. For preparative scale separation of DNA fragments 100 to 200 μg of restricted DNA was applied to a standard gel in a 10 x 0.5 x 0.3 cm slot.

2.10.2 Polyacrylamide gels

Polyacrylamide gels (PAA gels) are formed from acrylamide, a small organic molecule with an aminocarbonyl group ($-CONH_2$) and bisacrylamide, which consists of two acrylamide monomers linked through their aminocarbonyl groups. The cross-linking of these two components to form the gel matrix is catalyzed by ammonium persulfate and tetramethylene diamine (TEMED). The ratio of the concentration of acrylamide to bisacrylamide

determines the density of the crosslinking: the higher the ratio, the looser the network.

2.10.2.1 Non-denaturing polyacrylamide gels

30% (w/v) stock solutions with three different ratios of acrylamide:bisacrylamide (50:1, 30:1, 19:1) were prepared after cleaning and deionizing the acrylamides by mixing with 2 gm of activated charcoal/liter and 10 gm of Amberlite MB1, filtering through a glass fiber filter and then a 0.45 um millipore filter (Type HAWP). For analytical and preparative separation of DNA fragments smaller than 1 kb, 5 to 10% PAA gels with 1 mm thickness, 20 cm width and 20 to 40 cm length were poured between two glass plates and run in a vertical position using TBE as running buffer. For each gel, an aliquot of the stock solution was diluted with 1X TBE and polymerization effected by the addition of AP to 0.5% and 8 x 10^{-4} v of TEMED. The gel was poured between two acetone-cleaned glass plates and a comb with 1 cm wide slots, or for preparative separations 4 cm wide slots, inserted. After a 30 to 60 min polymerization, the comb was removed. The 20 μl samples were supplemented with 2 μl of 10X TBE and 2 μl of PAA sample buffer then loaded with a Hamilton syringe, and electrophoresed with a constant current of 35 to 40 mAmp. 5% gels were used to separate fragments in the range 10 to 1000 bp and 10% gels for fragments from 10 to 300 bp. A Hae III digest of ØX174 or a Hinf I digest of pBR322 were used as length standards.

For strand-separation of DNA fragments labeled at both 3'ends, preparative to DNA sequencing, 5% gels with an acrylamide:bisacrylamide ratio of 50:1 were cast using 20 or 40 cm plates and the 4 cm comb. The DNA sample was dissolved in a fresh solution containing 30% v/v DMSO, 1 mM EDTA, 0.05% Xylene cyanol and 0.05% Bromophenol blue. The sample was heated in a 90°C water bath for 2 min to denature the DNA, then quickly loaded; gels were run at 8 V/cm until the double-stranded form had reached the bottom of the gel as judged by the positions of the dye markers. A test strand-separation was always run first, as the single strands of a particular fragment do not necessarily separate as needed.

PAA gels were stained in 1X TBE containing 0.5 µg/ml EtBr for 15 min. Gels in which radiolabeled fragments had been separated were exposed to X-ray film (Kodak AR or Fuji) with Dupont Quanta III intensifying screens at -70°C. Photographs of agarose or polyacrylamide gels were made using transmitted UV light of 254 or 310 nm, a Polaroid MP4 camera with Polaroid film Type 667 or 665 and a red light filter (No. 3). With a lens setting of 4.5, exposure times were between 3 and 30 s depending on the amount of EtBr-complexed DNA present per band.

2.10.2.2 Denaturing polyacrylamide gels

Denaturing polyacrylamide gels containing 7M urea were used in S1 nuclease analysis and to separate fragments after DNA sequencing reactions. Their preparation except for the inclusion of 7M urea is the same as for non-denaturing PAA gels. Since sequencing gels were 40 to 80 cm long and 20 cm wide but only 0.3 cm thick, it was found advantageous to pour the gel on the lower plate then slowly lower the upper plate down on top of it and seal the mold by clamping along the sides. The comb for this gel contained slots having approximately 0.4 to 0.5 cm width. If the gel was to be used the next day, it was sealed above the slots with polymerized PAA.

The samples to be loaded on denaturing gels were dissolved in small amounts of DNA sequencing sample buffer and heated to 90°C for 2 min before loading. Between 0.5 and 1.0 µl of sample could be loaded per slot through the use of a 0.1 mm thick glass capillary. The 80 cm gels were usually cast with a 6% PAA solution (19:1) thereby allowing the separation of fragments between 30 and 150 bp with one loading and fragments from 120 to 350 bp with a second loading. Gels were run at 10 to 40 mAmp, 2000 to 5000 V, using 1X TBE as running buffer. The progress of the separation could be monitored with the dye markers. On a 6% gel, the bromophenol blue runs around 30 bp and the xylene cyanol at about 90 to 100 bp. 20% PAA gels (19:1) cast on 40 cm plates were used to separate fragments having sizes from 3 to 50 bp.

2.10.3 RNA electrophoresis in formaldehyde/agarose gels

The polyA-RNA fraction of spinach leaf RNA recovered after oligo-dT

affinity chromatography (Westhoff et al., 1981) was fractionated under denaturing conditions by electrophoresis in agarose gels containing 2.2 M formaldehyde (Lehrach et al., 1977). The electrophoresis was buffered with MOPS. A 1% gel was prepared by dissolving 2.1 gm of agarose in 217.5 ml water, autoclaving for 35 min, cooling to 60°C and adding 52.5 ml formaldehyde (37%) and 30 ml of 10X MOPS. The gel solution was poured on a 20 x 20 cm glass plate that had been sterilized by dry heat (200°C) overnight. The comb, which was sterilized by irradiation with ultraviolet light (365 nm, overnight) formed wells with dimensions of 1.0 x 1.0 x 0.5 cm for analytical purposes and 4.0 x 1.0 x 0.5 cm for preparative purposes. After hardening, the gels were covered with saran wrap and allowed to set at least one hour before loading the samples. Analytical samples contained 25 µg of RNA, 5 µl of 10X MOPS, 8.75 µl of 37% formaldehyde, and 25 µl of 100% deionized formamide in a total of 50 µl. Preparative samples contained 100 to 200 µg of RNA in a total of 100 µl. The samples were incubated at 55°C for 15 min to denature the RNA without degrading it, placed on ice, and one-tenth volume agarose sample buffer added before loading. Electrophoresis was carried out under a fume hood at 50 mAmp until the bromophenol dye marker had traversed 80-90% of the 20 cm gel. The marker RNA (20 µg) was stained for 60 min in 0.1 M NH_4-acetate before photographing.

2.11 DNA fragment preparation

2.11.1 Electroelution of DNA from agarose or PAA gels

After separation of DNA fragments by electrophoresis through an agarose or PAA gel of appropriate concentration, the individual fragments, which could be visualized under UV light due to their intercalated EtBr, were cut out with a sharp razor blade and placed separately in a bag of pre-boiled dialysis tubing (Visking size 2-18/32", Medicell International, LTD.) containing TE. The DNA was forced to wander to the wall of the dialysis bag under a current of 70 mAmp, the current direction reversed for 20 sec to detach the DNA from the membrane and the DNA collected in a siliconized 30 ml Corex tube. Contaminants were removed by successive extractions with phenol and chloroform. The aqueous phase was concentrated by extraction with

n-butanol (Maniatis et al., 1982) and the DNA precipitated. DNA recovered by this method could be contaminated with sulfated polysaccharides which inhibit ligases, polymerases and restriction enzymes. The DNA can be further purified by passage through an Elutip column (Section 2.11.3).

2.11.2 Salt extraction of DNA from PAA gels

Radiolabeled and unlabeled DNA fragments were recovered from PAA gels by a modification of the method of Maxam and Gilbert (1980). Gels were stained with EtBr, the desired fragments cut out and eluted overnight in salt extraction buffer at 37°C. For fragments over 500 bp in length, the PAA gel material was pulverized with a siliconized glass rod before addition of buffer. After overnight incubation, the aqueous phase was removed, its volume measured and the NaCl concentration adjusted to 0.2 M for cleaning over Elutip.

2.11.3 Purification of DNA by column chromatography

Two types of column chromatography were used to purify DNA:

(1) Molecular exclusion chromatography

The column, poured in a pasteur pipette plugged with siliconized glass wool, was formed from either Sephadex G-100 (molecular exclusion limit for globular molecules 230 bp) or Sephacryl S200 Superfine grade (molecular exclusion limit 380 bp). The column was equilibrated with TE, the sample DNA applied in a small volumne (10 to 20 µl) to the top of the column, allowed to flow in then eluted with TE. DNA is too large to enter the crosslinked Dextran bead matrix and emerges with the void volume. The progress of the chromatographic separation can be followed optically be addition of appropriate dye markers or if the DNA is radiolabeled with a Geiger counter. This method is rapid and suitable for the separation of radiolabeled DNA from unincorporated nucleotides after nick translation or end-labeling.

(2) Absorption chromatography

In this method, the DNA is selectively bound to the column material (Elutip, Schleicher and Schull) at a low NaCl concentration (200 mM), contaminants are washed away and the DNA released by elution with 1.0 M

NaCl. This method yields a very high purity DNA and a dilute DNA solution can be concentrated, however, a certain fraction of the DNA remains irreversibly bound to the column, for Elutip, 5 to 10%. Elutip also employs a filter which removes large debris and the column can be regenerated wtih low salt.

2.12 Hybridization techniques

2.12.1 Southern and Northern transfers

The technique for the transfer of DNA fragments from agarose gels to nitrocellulose filters such that the original pattern of fragment separation is preserved was developed by Southern (1975). DNA fragments are transferred in proportion to the log 10 of their molecular weight. A more quantitative transfer can be achieved by first soaking the gel in 0.25 M HCl (Wahl et al., 1979) causing partial depurination of the DNA. The gel is then rinsed briefly with distilled water and soaked in the denaturing solution (0.5 M NaOH, 1.5 M NaCl) for 30 min. During this step DNA denatures and strand cleavage occurs at sites of depurination. The gel is then neutralized in 0.5 M Tris-HCl, pH 7.0, 3.0 M NaCl for 30 min and placed on top of several sheets of Whatman 3 MM paper (of the same width as the gel) whose ends then act as wicks by soaking in 16XSSC (2.4 M NaCl, 0.24 M Na-acetate). Plastic strips are placed around the gel to disallow sideways diffusion of buffer. The nitrocellulose paper, cut exactly to the dimensions of the gel and wetted by careful floating on 2XSSC, is placed on top of the gel and air bubbles carefully removed. The nitrocellulose is covered with a second sheet of 3 MM paper, a 5 cm stack of paper is added and a glass plate to evenly distribute the weight. Most of the DNA under 1 kb in size is transferred in a few hours by the upward flow of the salt solution. After transfer, the filter is soaked in 2XSSC for 15 minutes before baking for 2 hours at 80 °C. These filters can be stored dry for several months before use.

The transfer of RNA to nitrocellulose from formaldehyde/agarose gels is essentially the same. After electrophoresis the gel (without prior treatment) is placed on 2 sheets of 3 MM paper saturated with 20XSSC (3.0 M NaCl, 0.3 M Na-Citrate), the nitrocellulose (wetted with

2XSSC laid on top, covered with 2 additional sheets of 3 MM paper and a 5 cm stack of paper towels as before. After a 12 to 15 hr transfer, the Northern blot is air-dried then baked for 2 hrs at 80°C. Washing the blot before baking will remove most of the RNA. This procedure is adapted from Thomas (1980).

The nature of the binding of nucleic acids to nitrocellulose is not known although it is known that it is weakened by a rise in temperature and strengthened by a rise in salt concentration. Small DNA fragments (less than 250 bp) are not bound efficiently.

2.12.2 Radiolabeling of DNA

Three procedures were used for radiolabeling DNA: nick-translation (Rigby et al., 1977; Maniatis et al., 1975), 3' end-labeling (Maniatis et al., 1982) and 5' end-labeling (Maxam and Gilbert, 1980).

(1) Nick-translation

E. coli DNA polymerase I, the main enzyme responsible for bacterial chromosome replication and repair, has two activities useful for radiolabelling DNA: a 5' to 3' exonucleolytic activity which removes nucleotides sequentially from the 5' side of a nick and a 3' to 5' polymerizing activity which adds nucleotides sequentially to the 3' side of a nick. Nick translations were performed in a minimal volume (10 to 15 µl) since the enzyme is more efficient with dNTP concentrations above 2 µM. The assay contained 1 µg DNA, 50 mM Tris-HCl, pH 7.2, 10 mM Mg_2SO_4, 0.1 mM DTT, 100 µmoles each of unlabeled deoxynucleotidetriphosphates, 10 to 20 µCi of (α-^{32}P)dATP or dCTP (0.4-0.6 µM) and 1 or 2 units of DNA polymerase I. Incubation was allowed to continue for 1 to 3 hrs at 16°C. Labeled DNA was separated from unincorporated nucleotides by column chromatography (Sephacryl S200) or by selective precipitation with 2M NH_4-acetate and 3 v of ethanol. The size of DNA after nick-translation depends on the amount of DNase I present during the reaction. Differing batches of DNA polymerase I contain different amounts of DNase I. The best concentration results in about 30% incorporation of labeled nucleotide into the DNA and the DNA strand length varies between 400 to 800 bp.

(2) 3' End-labeling

The Klenow fragment of DNA polymerase I (Jacobsen et al., 1974) carries the 5' to 3' polymerizing activity of the holoenzyme. 3' recessed termini created by restriction endonuclease digestion are filled most readily but blunt-ended fragments can be labeled by replacement of the nucleotide present at the 3' hydroxl terminus. DNA termini with 3' extensions are not labeled efficiently. Labeling at 3' ends was performed with (α-^{32}P) dNTPs in 10 μl of 50 mM Tris-HCl, pH 7.2, 10 mM $MgSO_4$, 1 mM DTT and 2.5 units of the Klenow fragment at 20°C for 30 to 60 min. When the first base to be filled in was not an A or C, the appropriate dNTP was added at a concentration of 100 μM. The reaction was stopped by a 5 min incubation at 65°C and labeled DNA separated from unincorporated nucleotides by column chromatography (Section 2.11.3).

(3) 5' End-labeling

To map the 5' end of an mRNA, the terminal 5' phosphates of the appropriate DNA fragment were first removed by treatment with calfs intestinal phosphatase (CIP). 20 pmole of 5' phosphates were removed in a 50 μl assay containing 50 mM Tris-HCl, pH 9.0, 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$, 1 mM spermidine, with 5 units of CIP enzyme by a 30 min incubation at 37°C. To inactivate the enzyme, the reaction volume was increased to 100 μl containing 35 mM Tris-HCl, pH 8.5, 100 mM NaCl, 1 mM EDTA, 0.5% SDS and heated to 68°C for 15 min. The enzyme was removed by two successive extractions with phenol-chloroform (1:1) and two extractions with chloroform. The DNA was precipitated and resuspended in 50 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 5 mM DTT, 0.1 mM spermidine, 0.1 mM EDTA, to which (γ-^{32}P) ATP (specific activity 3000 Ci/mmole) was added to a concentration of 20 μmolar and 1 unit of T4 polynucleotide kinase for a total reaction volume of 10 μl. The DNA was radiolabeled as the kinase phosphorylates the 5' hydroxyl group during a 1 hr incubation at 37°C. The enzyme was removed by phenol:chloroform (1:1) extraction and radiolabeled DNA separated from unincorporated nucleotides by Sephacryl S200 column chromatography (Section 2.11.3).

2.12.3 DNA hybridization to nitrocellulose-bound nucleic acids

In DNA-DNA or DNA-RNA hybridization reactions two single-stranded molecules anneal to from a base-paired duplex. The highest rate of nucleic acid hybridization occurs approximately 25°C below the temperature (T_m) at which the DNA is 50% melted. T_m varies with G + C content of the nucleic acid and with the fidelity of base pairing. Four useful relationships must be considered when choosing hybridization and washing conditions:

(1) T_m= 69.3 + 0.41 (G + C%) (Marmur and Doty, 1962).

(2) $(T_m)_{u2}$- $(T_m)_{u1}$= 18.5 log (u2/u1), where u1 and u2 are the ionic strengths of two solutions (Dove and Davidson, 1982).

(3) A 1% increase in the number of mismatched base pairs decreases the T_m by 1% (Bonner et al., 1973).

(4) Every 1% increase in formamide concentration reduces the T_m by 0.7°C (McConaughty et al., 1969; Hutton, 1977).

For spinach chloroplast DNA with a G + C content of 38% (Herrmann et al., 1975) its T_m in 1XSSC (Eq. 1) is 85°C. In 5XSSC (Na+= 0.82 M) the T_m is raised 12.8°C (Eq. 2) to 97.8°C. Under washing conditions of 2XSSC (Na+= 0.33 M) the T_m is raised 5.5°C to 90.5° C. However since both hybridization and washing were carried out in 50% formamide, the T_m decreases in each case by a factor of 0.7(50)= 35°C.

Hybridizations were carried out in 50% formamide, 5XSSC at 42° C which is T_m-21°C and washing in 50% formamide, 2XSSC at 42°C which is T_m-13.5°C. The rate of DNA-DNA hybridization in 50% formamide is about two times slower than in aqueous solutions (Casey and Davidson, 1977). However, the kinetics of hybridization are difficult to predict since both the exact concentration of the immobilized nucleic acid and its hybridization availability are unknowns.

The hybridization mixture (DNA-DNA) also contained 5X Denhardt's solution (500 mg Ficoll, 500 mg polyvinylpyrrolidone, 500 mg bovine serum albumin) and 100 μg/ml calf thymus DNA sheared to approximately 400 bp and heated to 90°C before usage. Denhardt's is a solution of high molecular weight polymers which prevents further binding of single-

stranded DNA to the nitrocellulose, but does not affect the specific annealing of complementary sequences. The hybridization is promoted by a high concentration of salt which masks the mutual repulsion between two highly charged DNA molecules.

Filters were pretreated by floating them DNA or RNA side up on 5XSSC and then submerging them for 5 min. They were then incubated overnight sealed in a plastic bag containing the prehybridization solution: 50% formamide, 5XSSC, 5X Denhardt's, 100 μg/ml sheared calfs thymus or salmon sperm DNA (denatured by boiling 10 min before addition) using 0.2 ml solution/cm^2 filter. On the following day, the prehybridization solution was discarded and replaced with the hybridization solution containing the same ingredients and heat-denatured ^{32}P-labeled DNA. However, the volume of the hybridization solution was reduced to 50 to 100 μl/cm^2 of filter. Incubation was at 42°C overnight for DNA-DNA and for 50 hours for DNA-RNA hybridizations. Hybridizations to Northern blots included 0.1% SDS. After the appropriate hybridization period, the hybridization solution was replaced with 100 ml of washing solution containing 50% formamide, 2XSSC, and 0.1% SDS and incubated at 42°C for 30 to 60 min. This wash was followed by three successive washes with 200 to 300 ml each of 2XSSC, 0.1% SDS at 42°C to remove any traces of formamide. The wet filters were sealed in plastic and exposed to X-ray film with a Quanta III intensifying screen.

2.12.4 Grunstein-Hogness colony screening

After the detection of transformant bacteria containing foreign DNA, further screening is often necessary to find the recombinant plasmid(s) containing a particular fragment of DNA. The procedure developed by Grunstein and Hogness (1975) allows the rapid screening of several thousand recombinant plasmids by hybridization of their DNA with an appropriate radioactive RNA or DNA probe.

The recombinant bacteria are transferred separately with a sterile toothpick (or by replica plating) to a nitrocellulose filter placed on the surface of an agar plate (appropriately supplemented with antibiotics) and to a second agar plate of the same media (Master plate), which

will later provide the source for further growth of the desired plasmid(s). The plates are incubated at 37°C upside down until the streaks on the nitrocellulose reach about 1 mm width (approximately 8 hrs). The master plates are incubated overnight then sealed with parafilm and stored at 4°C. The nitrocellulose is lifted from the agar with tweezers and placed bacteria side up, successively for 5 min each, on Whatman 3 MM paper saturated with the following solutions: (1) 10% SDS, (2) 0.5 M NaOH, 1.5 M NaCl, (3) 1.5 M NaCl, 0.5 M Tris-HCl, pH 7.4, (4) 2XSSPE. Between each transfer, the filter is briefly blotted dry from below. The filters are allowed to air dry and are then baked at 80°C for 2 hours. SDS, which is optional, appears to give a sharper hybridization signal perhaps by limiting the diffusion of the plasmid DNA during lysis and denaturation with NaOH and subsequent neutralization. Colonies can also be grown on Millipore HAWP filters or on Whatman 541 paper in which case step (4) is followed by rinsing with 95% ethanol.

2.13.1 DNA sequencing by base-specific chemical cleavage

In the method developed by Maxam and Gilbert (1977, 1980) a terminally labeled double or single-stranded DNA molecule is partially cleaved at each of the four bases in four separate reactions, and the products resolved by size on a denaturing polyacryamide gel. The chemical cleavage involves three consecutive steps: modification of a base, removal of the modified base from its sugar, and strand scission at that sugar. Specificity lies in the first reaction, which occurs with only a few percent of the bases, whereas the second and third reactions must be quantitative. It is therefore very important that the temperature and duration of the first reaction are carefully controlled.

Base modification reactions:

(1) Cleavage at pyrimidines (CT and C reactions)

Hydrazine, NH_2NH_2, is a nucleophile which reacts with the bases cytosine and thymidine by opening the ring between the 5,6 double bond and leaving a new five-membered ribosylurea ring. Further reaction with hydrazine releases this ring and the urea fragment on the deoxyribose,

leaving the sugar in the DNA backbone as a hydrazone. Partial hydrolysis is effected by 15-18 M hydrazine at 20°C. In the presence of 2M sodium, the reaction with thymine is preferentially suppressed. Thus a comparison of the reaction with salt and without salt reveals where the chain was broken by a C and where by a T, respectively. 5' methyl cytosine, a methylated analog of cytosine, may be created by the dcm methylase of E. coli in the sequence $C^{me}CAGG$ or $C^{me}CTGG$. The methyl group interferes with the action of hydrazine and the lack of cleavage appears as a gap in the sequence ladder. Nevertheless, the opposite strand must contain a guanine.

(2) Cleavage at purines

a) Cleavage at Guanine (G reaction)

Dimethyl sulfate (DMS) methylates guanines in DNA at the N7 position and adenines at the N3 position. Guanines methylate 5 times faster than adenines. The glycosidic bond of a methylated purine is unstable at pH 8.0 and breaks displacing the base and leaving the open sugar ring.

b) Cleavage at adenine and guanine (AG reaction)

An acid hydrolysis reaction is used to detect the positions of the purines adenine and guanine. A partial hydrolysis is effected by using 66% formic acid since the rate of the reaction is a linear function of the concentration of protons. Purines are much more readily hydrolyzed than pyrimidines because they are able to stabilize the Schiff base intermediate, but cytosine will also react. Comparison of the G and the AG reactions indicates where the chain was broken by a G and where by an A.

c) Adenine and cytosine cleavage (A>C reaction)

This additional reaction was useful for identifying adenines when thymidines were not properly suppressed in the C reaction. Under alkali conditions (1.2 N NaOH) adenine rings (85%) and to a lesser extent cytosine rings (15%) open adjacent to the glycosidic bond.

Strand scission by piperidine:

After the reagents for the base-modification reactions have been removed by precipitation and washing with ethanol, the dried samples are redissolved in 1.0 M piperidine. Piperidine opens the 7-methyl guanines

in the G reaction, displaces all ring-opened bases from their sugars (CT, C, C>A and AG reactions) and catalyzes the ß-elimination of phosphates from the empty sugars to cleave the DNA backbone. This reaction is carried out at 90°C for 15 to 30 min.

Freshness and purity of the sequencing reagents is very important. Anhydrous dimethyl sulfate may absorb moisture from the air and hydrolyze to methanol and sulfuric acid. Hydrazine is readily oxidized to the diimine and piperidine can also undergo oxidation. Reagents of high purity must be used and kept tightly sealed from the atmosphere or under N_2. Oxidized hydrazine, in particular, does not allow salt suppression of thymines in the C reaction. DMS and hydrazine are poisonous and volatile and must be inactivated in small portions with 5 M NaOH and 3 M ferric chloride, respectively. The 3' labeled DNA was dissolved in 6 to 8 µl of a 1 mg/ml solution of sheared calfs thymus DNA in water. Other reagents were kept on ice. The steps in the sequencing reactions are listed in Table 1. The base modification reactions were allowed to proceed for various times depending on the length of the fragment being sequenced. After modification the DNA was precipitated, recovered by centrifugation, reprecipitated with 0.3 M Na-acetate and 3 v of ethanol, washed twice with 70% ethanol and dried under vacuum. After strand scission with piperidine the DNA was precipitated, collected, dried, counted by the method of Cerenkov and resuspended in DNA sequencing sample buffer such that the C and G reactions contained 5-10,000 cpm/µl and the CT, AG and AC contained 10-20,000 cpm/µl.

2.13.2 Computer analysis of DNA sequences

Several computer programs written by Kröger and Kröger-Block (1982) were used to analyze the DNA sequence data. These included programs for the identification of restriction endonuclease sites (P1), for the translation of both strands of the DNA sequence in all posssible reading frames, for the comparison of homologous DNA sequences (SUCHSEQUENCE), and for the calculation of the codon usage and molecular weight of a particular protein or open reading frame. A program for the production of a hydropathy profile according to the method of Kyte and Doolittle (1982) was kindly written by Dr. Otto Wasmund. In this method, each

REACTION TYPE	C	CT	AG	G	AC
^{32}P-LABELED DNA	1	2	2	1	2
5 M NACL	20	-	-	-	-
DISTILLED WATER	-	18	2	-	-
30 M HYDRAZINE	30	30	-	-	-
100% FORMIC ACID	-	-	8	-	-
DMS BUFFER	-	-	-	200	-
DMS	-	-	-	1	-
1 N NAOH	-	-	-	-	20
LENGTH (BP)	20°C INCUBATION TIME (MIN)				
LARGER THAN 1000	14	14	4	8	8*
300-1000	16	16	5	9	8
200-300	20	20	6	12	9
100-200	25	25	9	15	10
LESS THAN 100	30	27	11	18	15

Cool 10 min in methanol/dry ice bath (-70°C).
Pellet in Eppendorf centrifuge 15 min (15,000g).
Wash once with 70% ethanol.
Dry under vacuum 15 min.
Resuspend in 250 µl 0.3 M Na-acretate, 750 µl 100% ethanol.
Cool -70°C, pellet, wash, dry as before.
Resuspend in 100 µl 1 M piperidine. Incubate 90°C, 20 min.
Add 10 µl 3 M Na-acetate, pH 6.0; 300 µl 100% ethanol.
Precipitate overnight at -20° C.
Pellet in Eppendorf, wash and dry.
Resuspend in DNA sequencing sample buffer.

Table 1: Base-specific chemical reactions for DNA sequencing. Volumes are in microliters. * The AC reaction incubation is at 90° C.

amino acid is rated on a hydropathy scale, which ranges from 4.5 for isoleucine, judged to be the most hydrophobic residue, to -4.5 for arginine, the most hydrophilic residue. The average hydropathy value of a moving segment (usually chosen to include 9 to 15 residues) is calculated over the entire sequence and consecutive scores are plotted from the amino to the carboxyl terminus. For membrane bound proteins, portions of their sequences located within the lipid bilayer are clearly seen as large uninterrupted areas on the hydrophobic side of the midpoint line. The hydropathic values for the individual amino acids were established by taking into account water-vapor transfer free energies (Wolfenden et al., 1979) and interior-exterior distribution of amino acid side chains as determined by Chothia (1976).

2.13.3 RNA secondary structure

Stability of secondary RNA structures was calculated according to Tinoco (1973). Negative free energies are assigned to base pairs considered two at a time and positive energies added for the three kinds of loops: interior, bulge, and hairpin.

2.14 S1 nuclease analysis

The main application of nuclease S1 is to trim single-stranded protruding ends of DNA or RNA without significant nibbling at duplex ends. The mapping of 5' and 3' ends of mRNA can be accomplished by analysing S1 resistant RNA-DNA hybrids in denaturing PAA gels. S1 nuclease analysis may also yield information about the secondary structure of mRNA by digestion at hairpin loop structures in an otherwise base-paired RNA-DNA hybrid. 5' and 3' end-labeling of appropriate fragments were performed as described in Section 2.12.2.

The RNA-DNA hybridization was performed in 30 µl containing 80% formamide, 30 µg plastid RNA, 400 mM NaCl, 40 mM Pipes-NaOH, pH 6.8, 1 mM EDTA and 1-2 pmole of ^{32}P-ATP labeled DNA. The assay was heated to 85°C for 15 min to denature the nucleic acids, then quickly cooled to 48°C and incubated for 4 hrs to allow RNA-DNA hybrids to form. A 48°C temperature is optimal for the formation of RNA-DNA hybrids in 80% formamide when the DNA has a G + C content of 40% (Casey and Davidson,

1976). In a control assay the RNA was either omitted or replaced with tRNA.

S1 nuclease was diluted to a concentration of 500 units in 200 µl of S1 nuclease buffer which had been preheated at 85°C for 15 min. Each sample received 500 units S1 nuclease in 200 µl of buffer and incubation was allowed to proceed for 30 min to 1 hr at 37°C. The samples were precipitated by the addition of 40 µg of tRNA and 3 v of ethanol. Two control assays were performed (1) DNA alone hybridization condition was submitted to S1 nuclease digestion, (2) DNA received neither hybridization nor S1 nuclease.

Controls and samples were resuspended in small amounts of DNA sequencing buffer, heated to 90°C for 5 min and loaded on 5 or 10% denaturing PAA gels.

3.0 Results

3.1 Location of genes for polypeptides of the reaction center core of the PS II complex

The gene for the P680 chlorophyll a apoprotein of the PS II reaction center (RC) has been located on the spinach plastid chromosome close to the genes for cytochrome b_6 and subunit 4 (Westhoff et al., 1983; Alt et al.,1983) some 40 kb away from the coding sequence for the 44 kd chlorophyll a binding apoprotein of the same RC. The 32 kd herbicide-binding protein is located almost directly opposite the P680 chlorophyll a apoprotein gene just outside the inverted repeat (Driesel et al., 1980). Figure 4 shows the locations of these three genes with respect to recognition sites for the endonucleases Sal I, Pst I, and Xho I. The gene for the P680 chlorophyll a apoprotein is located on the left half of Sal I fragment 7 which is contained in the recombinant plasmid pWHsp207 (see Section 2.6.2).

3.2 Sequencing strategy

In preparation for DNA sequencing specific restriction fragments were selected using the criteria (1) the entire plasmid could be digested and the resulting fragments cleanly separated and recovered, (2) restriction sites were so ordered such that single fragments could be selectively labeled at one 3' end using a specific combination of labeled and unlabeled nucleotides. Such primary fragments were used to establish most of the DNA sequence and to find secondary restriction enzyme sites which could be labeled to fill in the sequence gaps between primary fragments and on the opposite DNA strand. Secondary fragments usually contain the same 5' ends and therefore cannot be selectively labeled. When both 3' ends are labeled they must be separated by cutting the fragment with a new enzyme or by strand separation before sequencing.

3.3 P680 chlorophyll a apoprotein gene-restriction map

By cutting the purified pWHsp207 plasmid DNA with various restriction endonucleases (Eco RI, Hind III, Sal I, Bam HI, Hinc II, and Ava I) the following restriction map could be constructed for the Sal I fragment in the A orientation (Figure 5 and Wlaschek, 1982).

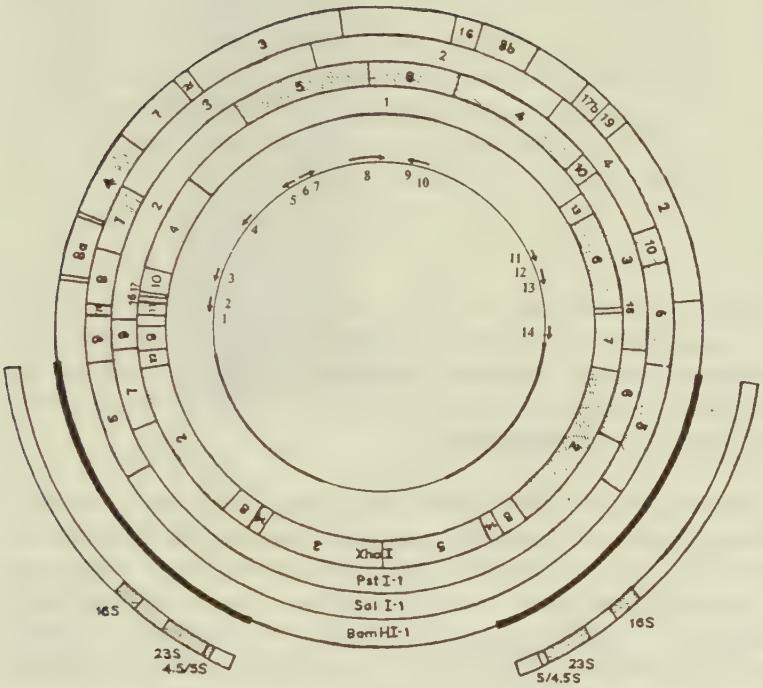


Figure 4: Map of spinach plastid chromosome showing recognition sites for Sal I, Pst I, Xho I and relevant Bam HI sites. Fragments are numbered in order of decreasing size. Extended sections indicate the inverted repeat of the chromosome. Locations of genes for thylakoid membrane proteins are indicated on the inner circle: (1) subunit 4, (2) cytochrome b6, (3) P680 chlorophyll a apoprotein, (4) cytochrome f, (5) large subunit ribulose carboxylase, (6,7) beta and epsilon subunits of CF_1, (8) two P700 chlorophyll a apoproteins, (9,10) 44 kd chlorophyll a apoprotein and D2 protein, (11,12) subunits I and III of CF_0, (13) alpha subunit of CF_1, (14) 32 kd herbicide-binding protein (from Herrmann et al., 1983).

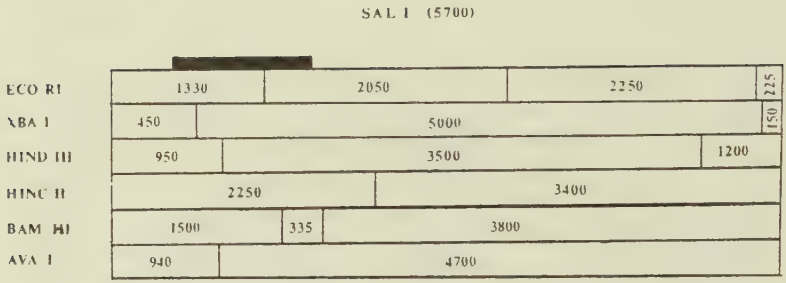


Figure 5: Map showing the positions of cutting sites for six restriction enzymes on Sal I fragment 7 of Spinacia oleracea plastid DNA. The position of the P680 chlorophyll a apoprotein gene is indicated by the thick black line (Morris and Herrmann, 1984). Numbers indicate the sizes of fragments in base pairs.

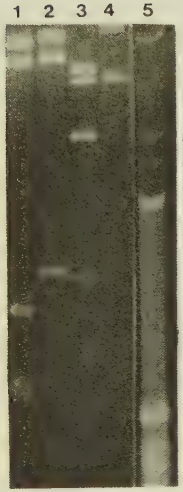
Knowing the primary restriction sites, double digestions could be performed producing DNA fragments conducive to sequencing. For this purpose, the plasmid with the Sal I fragment in the opposite orientation was also used (pWHsp 207b). Figure 6 shows the fragments obtained after digestion of pWHsp 207a with various enzymes and separation of the resulting fragments on an agarose gel. The fragment order is indicated in the accompanying scheme.

3.4 Fragment preparation for sequencing

A. SalI-Hind III fragments from pWHsp 207a

A double digest of the plasmid with Sal I-Hind III produces five fragments (Figure 6) of which the largest (4200 bp) has two Hind III ends, one in the P680 chlorophyll a apoprotein gene. The third largest fragment (960 bp) has one Hind III end within the gene and a Sal I end in the 3' flanking sequences to the gene. Fragments 1 and 3 were isolated from an agarose gel by electroelution and labeled at their 3' ends with the Klenow polymerase. Fragment 3 could be selectively labeled

A)



B)

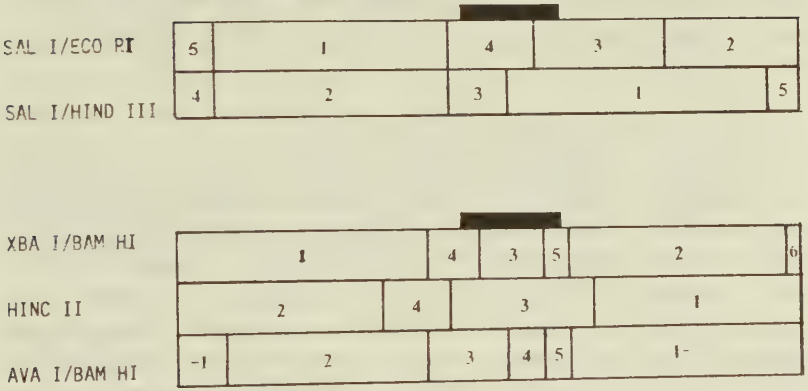


Figure 6: (A) Double digests of pWHsp 207a: lane (1) Sal I-Hind III, lane (2) Xba I-Bam HI; double digests of pWHsp 207b: lane (3) Hinc II, lane (4) Ava I-Bam HI, lane (5) Hind III digest of bacteriophage lambda and Hinf I digest of pBR322. DNA was separated on 0.8% agarose. (B) Diagram showing location of restriction fragments prepared for sequencing. Fragments are numbered in order of decreasing size. The location of the P680 chlorophyll a apoprotein gene is indicated by a thickened line.

at its Hind III end with α^{32}P-dATP and at its Sal I end with unlabeled dTTP and α^{32}P-dCTP. The two labeled ends of fragment 1 were divided by cleavage with Eco RI giving subfragments of 1500 and 450 bp which were separated on a PAA gel, recovered by salt extraction and sequenced.

B. Xba I-Bam HI fragments from pWHsp 207b

Double digestion of the plasmid in the B orientation with Xba I and Bam HI gives five fragments (Figure 6). Fragment 2 (3400 bp) has a Bam HI end in the 5' flanking sequences to the gene. Fragment 3 (1100 bp) an Xba I and a Bam HI end both within the gene. Fragment 4 (775 bp) has an Xba I end within the gene and fragment 5 (335 bp) has two Bam HI ends, one within the gene and one in the 5' flanking sequences. After electroelution from an agarose gel, fragments 3 and 4 were selectively labeled at their Xba I ends with α^{32}P-dCTP and Klenow polymerase and directly sequenced. Fragment 5 was labeled at both Bam HI ends with cold dGTP and α^{32}P-dATP and cleaved with Hinf I or Hpa II giving subfragments of 60/280 and 125/155 bp, respectively, which retained labeled ends. These fragments were recovered by salt extraction from a 10% PAA gel and sequenced.

C. Sal I-Eco RI fragments from pWHsp 207a

Double digestion of pWHsp 207a with Sal I and Eco RI produces six fragments (Figure 6). Fragment 3 (2050 bp) has one Eco RI end in the gene and the fourth fragment (1450 bp) has an Eco RI end within the gene and a Sal I end in the 3' flanking sequences. Fragments 3 and 4 were isolated by electroelution from a 1% agarose gel. Fragment 4 was labeled at the Eco RI site with α^{32}P-dATP and sequenced. Fragment 3 was labeled at both Eco RI ends, cut with Alu I and the resulting labeled 288 bp and 150 bp fragments isolated by salt extraction from a 5% PAA gel and sequenced.

D. Ava I-Bam HI fragments from pWHsp 207b

Double digestion of pWHsp 207b yields five fragments. The largest (4320 bp) contains a Bam HI site from the 5' flanking regions to the gene, the third fragment has an Ava I end in the gene, the fourth fragment contains the complementary Ava I end and a Bam HI end also

within the gene, and the fifth fragment has two Bam HI ends, one inside and one outside the gene. Fragments 1, 3, 4, and 5 were electroeluted from a 1% agarose gel. Fragment 1 was selectively labeled at the Bam HI end with unlabeled dGTP and α^{32}P-dATP; fragment 3 was selectively labeled at the Ava I end with unlabeled dTTP and α^{32}P-dCTP. Fragment 4 was labeled at both ends with unlabeled dTTP, dGTP, dATP, and α^{32}P-dCTP, the two ends separated by an Eco RI cut and the resulting 500 bp and 150 bp fragments recovered by salt extraction and sequenced.

E. Hinc II digest of pWHsp 207b

A Hinc II digest of pWHsp 207b gives four fragments. Fragment 3 (2250 bp) contains the entire P680 chlorophyll a apoprotein gene. It was therefore isolated in large quantities in order to prepare smaller fragments, e.g. Hinf I and Taq I.

F. Hinf I digest of Hinc II fragment III

The 2250 bp Hinc II fragment was cut with Hinf I into four fragments (Figure 7), which could be separated on a 3% agarose gel and electroeluted. All four fragments were labeled with α^{32}P-dATP. The strands of fragments 3 and 4 were separated on a 5% PAA gel, recovered by salt extraction and sequenced. Fragment 2, whose blunt Hinc II end was not labeled, was directly sequenced. The labeled ends of fragment 1 were separated by a Dde I cut and the 270 bp and 206 bp subfragments extracted from a 5% PAA gel and sequenced.

Figure 7: Hinf I subfragments of Hinc II fragment 3. The first row shows the sizes of the fragments in base pairs; in the second row the fragments are numbered in order of decreasing size. The thickened line indicates the position of the P680 chlorophyll a apoprotein gene.

G. Taq I digest of Hinc II fragment 3

A Taq I digest of Hinc II fragment 3 yields six fragments (Figure 8) which were separated and electroeluted using a 3% agarose gel. Fragment 2 was labeled at both ends with ^{32}P-dCTP, the labeled ends separated by a Hha I cut, the resulting 190 bp and 215 bp subfragments separated on a 5% PAA gel, salt extracted and sequenced.

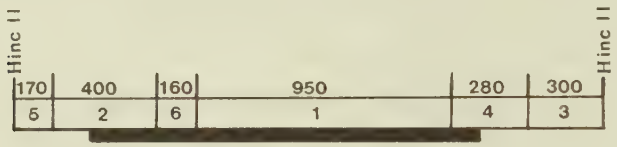


Figure 8: Taq I subfragments of Hinc II fragment 3. The first row shows the sizes of the fragments in base pairs; in the second row the fragments are numbered according to decreasing size. The position of the P680 chlorophyll a apoprotein gene is indicated by a thickened line.

3.5 DNA sequence of the P680 chlorophyll a apoprotein gene

An overview of the sequencing strategy for the P680 chlorophyll a gene is presented in Figure 9. The direction and extent of the twenty-two individual sequence analyses is indicated along with relevant restriction sites; in the accompanying Table 2 the fragments are listed individually and their treatment summarized.

In Figure 10 the nucleotide sequence of 2210 bp of spinach plastid DNA including the gene for the P680 chlorophyll a apoprotein and its flanking regions is given. Restriction endonuclease sites are located and the sequence is translated in all three possible reading frames on both strands.

In Figure 11, the sequence, beginning 300 bp proximal to the first methionine of the open-reading frame and ending at the junction between

Sal I fragments 7 and 8 is shown. The computer assisted translation in six possible reading patterns revealed only one large open-reading frame; it is capable of coding for a protein of 508 amino acids (M_r=56,246) starting from the first available ATG codon (methionine) at nucleotides 1-3 and ending with a TAA stop codon at position 1525-1527. The deduced amino acid sequence is shown below the nucleotide sequence. An arrow marks the putative transcription start site. The -10 and -35 regions of the putative promotor are boxed. The sequence AAGGGG located 8 to 13 bp proximal to the ATG start codon might function as a ribosome binding site (Shine and Dalgarno, 1974) by pairing with the sequence at the 3' end of plastid 16S rRNA during translation initiation. At positions -187 to -182 upstream from the first methionine codon, the sequence TAGAAT which shows similarity with the -10 bacterial promotor consensus sequence TATAAT (Pribnow, 1975) can be found. A corresponding -35 position sequence TTGCGT identifiable at nucleotide positions -211 and -206 also shows features in common with the E. coli consensus sequence TTGACA (Rosenberg and Court, 1979). A thirty-nine base pair stem-loop structure (Figure 12) located 295 bp distal to the TAA stop codon has a free energy of formation ΔG=-41.8 kcal calculated according to Tinoco et al (1973).

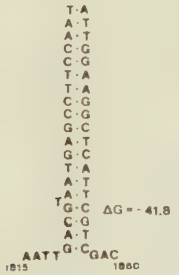


Figure 12: Proposed RNA stem-loop secondary structure in the 3' flanking sequences to the P680 chlorophyll a apoprotein gene. Numbering of nucleotides follows Figure 11. Free energy of formation, ΔG, was calculated according to Tinoco et al. (1973).

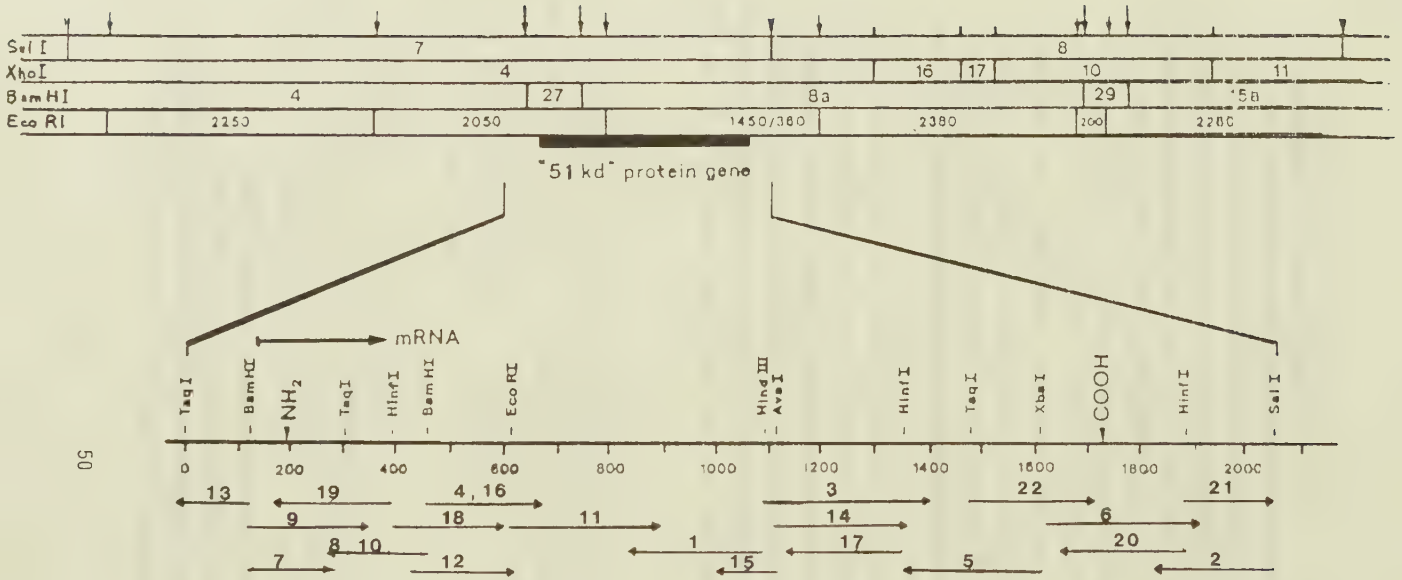
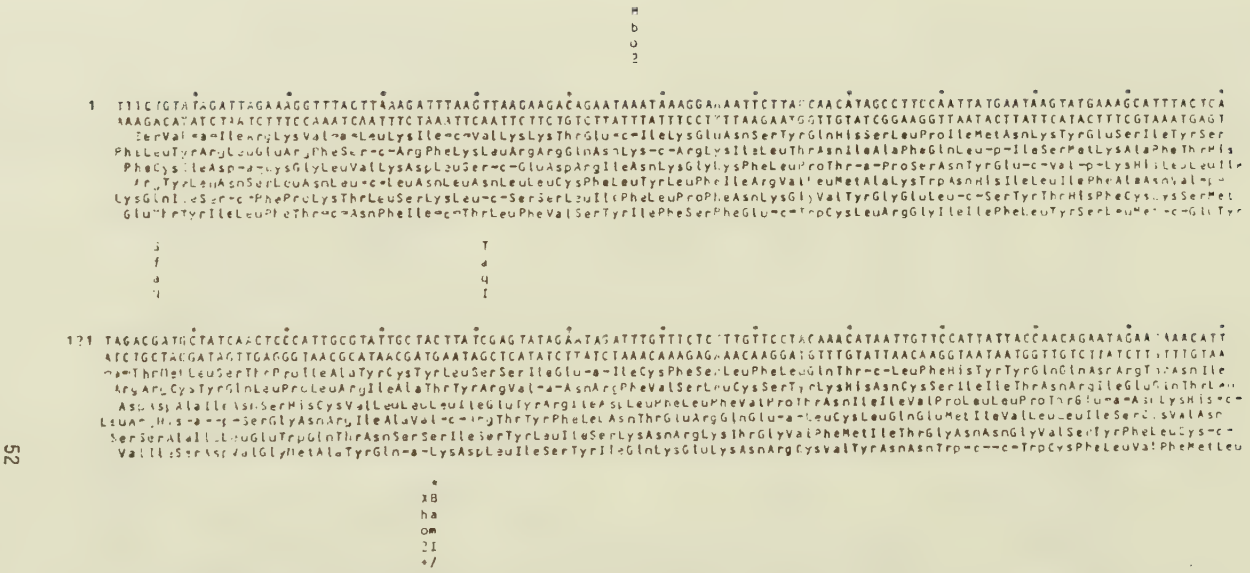


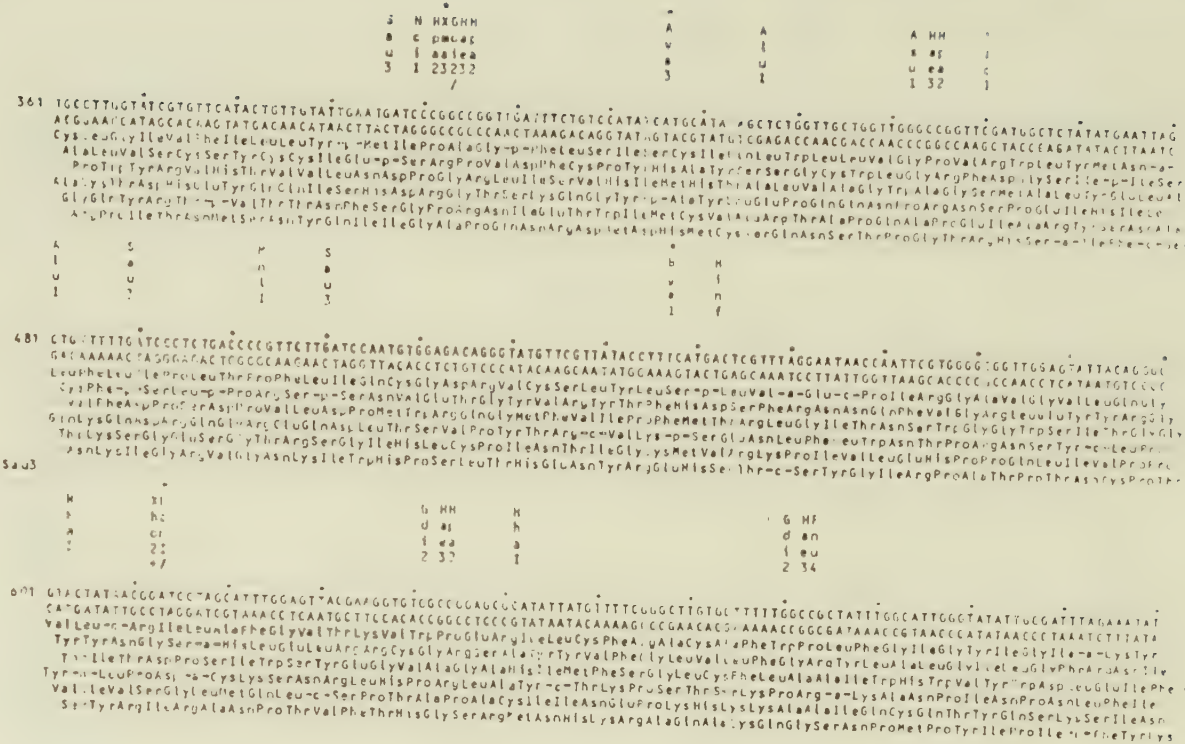
Figure 9: Upper part: Partial restriction map of fragments Sal I-7 and-8 of spinach plastid DNA. Fragment number is in order of decreasing size (Sal I, Xho I, Bam HI) or in base pairs (Eco RI). Thickened line indicates the location of the P680 chlorophyll a apoprotein gene. Lower part: Detailed map of the 2210 bp interval containing the apoprotein gene and flanking sequences. Thickened arrow indicates start and direction of transcription; small arrows illustrate direction and extent of individual DNA sequence analyses.

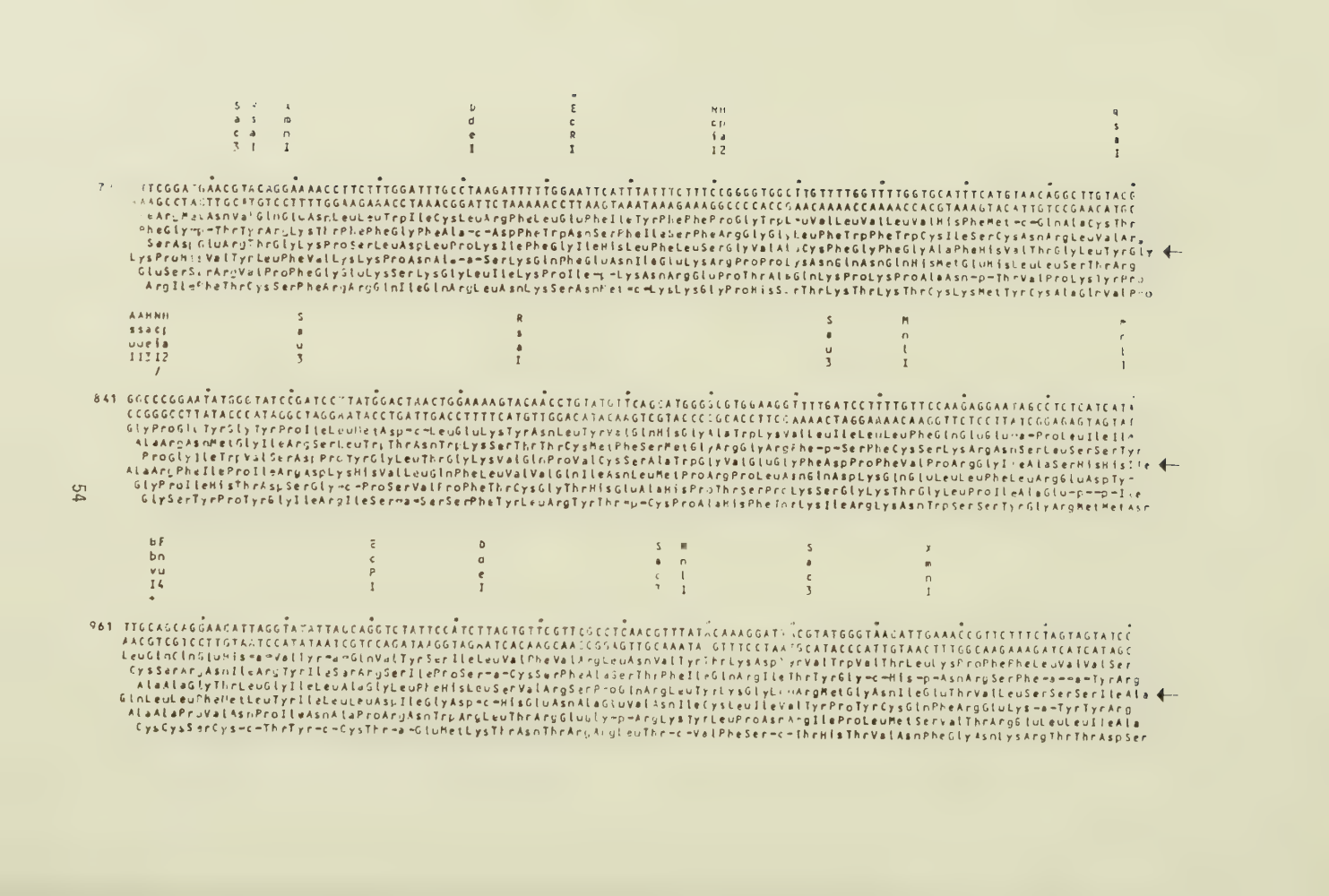
TABLE 2: DNA FRAGMENTS FOR SEQUENCE ANALYSIS OF THE P680 CHLOROPHYLL A APOPROTEIN GENE

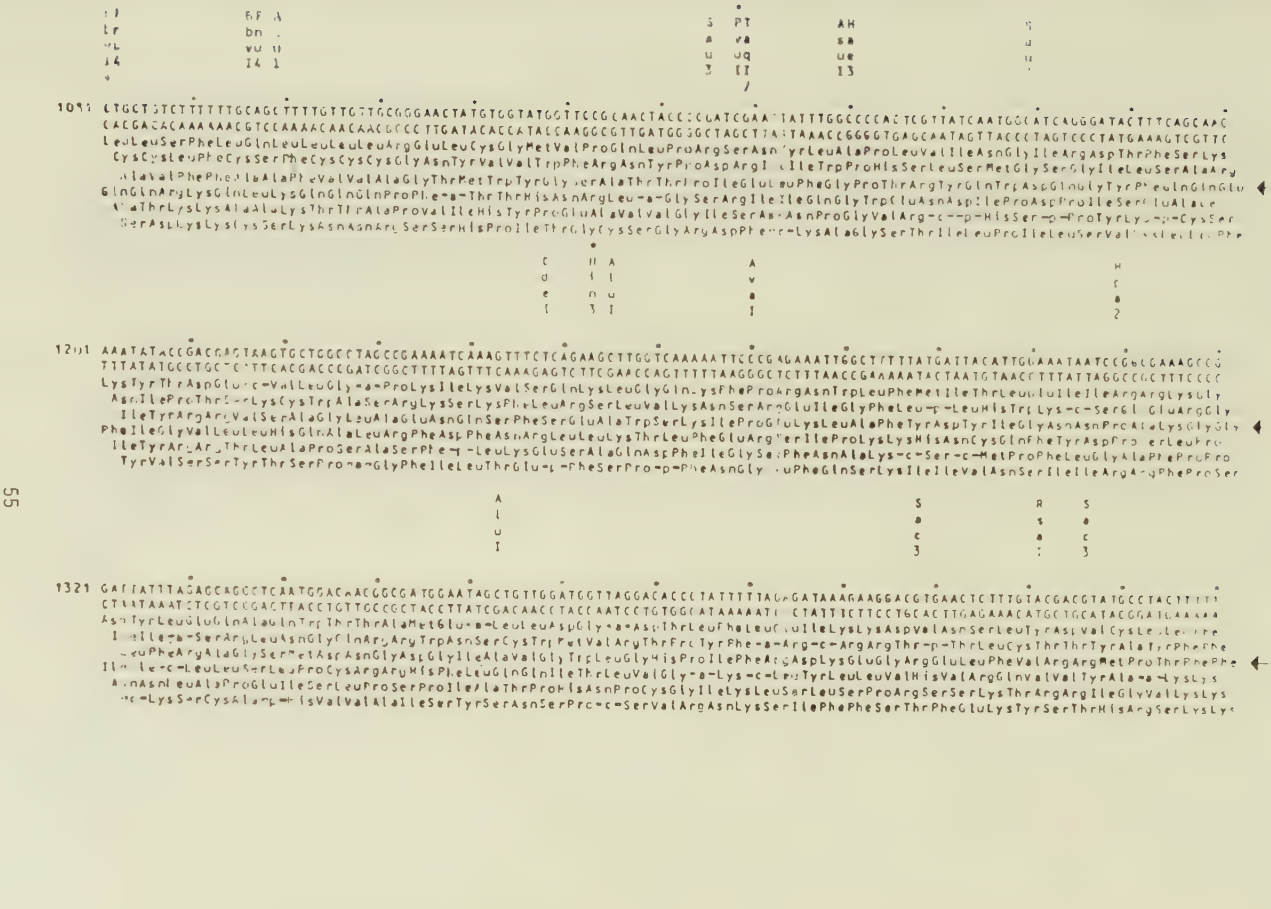
NO.	FRAGMENT	SIZE (BP)	LABELED END/SECOND CUT
1	Sal I-Hind III, 1	4200	Hind III/Eco RI
2	Sal I-Hind III, 3	960	Sal I
3	Sal I-Hind III, 3	960	Hind III
4	Xba I-Bam HI, 3	1100	Bam HI/Hind III
5	Xba I-Bam HI, 3	1100	Xba I
6	Xba I-Bam HI, 4	750	Xba I
7	Xba I-Bam HI, 5	335	Bam HI/Hpa II
8	Xba I-Bam HI, 5	335	Bam HI/Hpa II
9	Xba I-Bam HI, 5	335	Bam HI/Hinf I
10	Xba I-Bam HI, 5	335	Bam HI/Hinf I
11	Sal I-Eco RI, 4	1445	Eco RI
12	Sal I-Eco RI, 3	2050	Eco RI/Alu I
13	Ava I-Bam HI, 1	4320	Bam HI
14	Ava I-Bam HI, 3	1320	Ava I
15	Ava I-Bam HI, 4	668	Ava I/Eco RI
16	Ava I-Bam HI, 4	668	Bam HI/Eco RI
17	Hinc II/Hinf I, 1	900	Hinf I/Dde I
18	Hinc II/Hinf I, 1	900	Hinf I/Dde 1
19	Hinc 11/Hinf I, 2	700	Hinf I
20	Hinc II/Hinf I, 3	523	Hinf I/SS*
21	Hinc II/Hinf I, 4	170	Hinf I/SS
22	Hinc II/Taq I, 2	400	Taq I/Hha I

*SS= Single strand separation









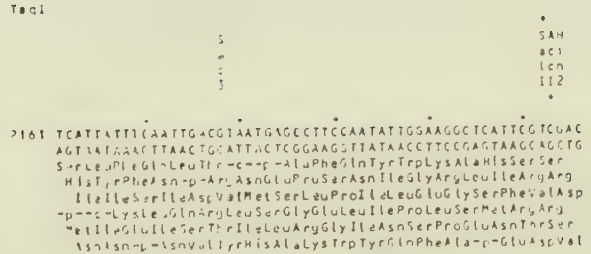
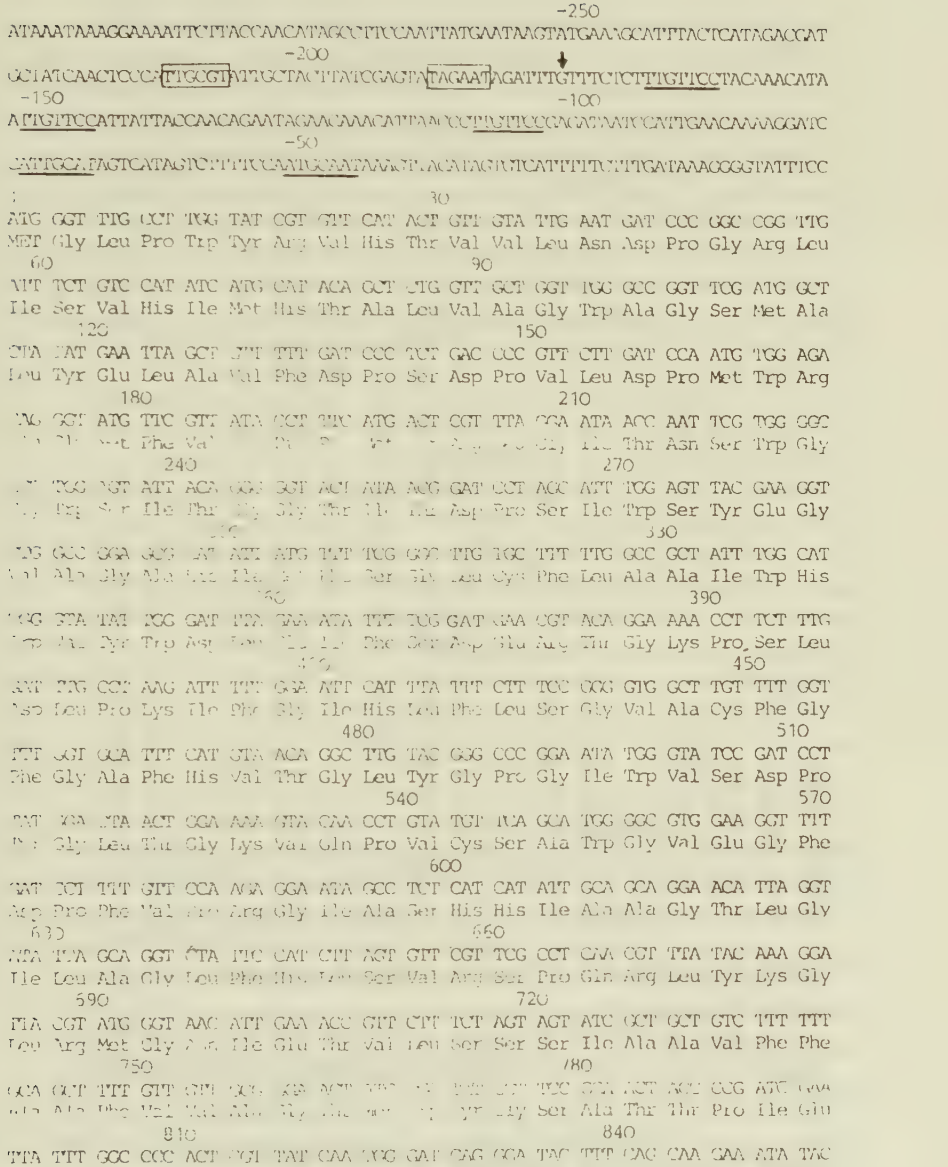


Figure 16: Nucleotide sequence of 2210 bp of spinach plastid DNA including the gene for the P680 chlorophyll a binding apoprotein of the PS II reaction center. The positions of selected restriction sites are shown and the sequence is translated in all six possible reading frames using the program described in Section 2.13.2. The position of the open reading frame coding for the P680 chlorophyll a apoprotein gene is marked with arrows.


-250
-200
-150
-100
-50
1
30
60
90
120
150
180
210
240
270
300
330
360
390
420
450
480
510
540
570
600
630
660
690
720
750
780
810
840

```
                                                                840
ACT CGT TAT CAA TGG GAT CAG GGA TAC TTT CAG CAA GAA ATA TAC CGA CGA GTA
Thr Arg Tyr Gln Trp Asp Gln Gly Tyr Phe Gln Gln Glu Ile Tyr Arg Arg Val
      870                                                       900
AGT GCT GGG CTA GCC GAA AAT CAA AGT TTC TCA GAA GCT TGG TCA AAA ATT CCC
Ser Ala Gly Leu Ala Glu Asn Gln Ser Phe Ser Glu Ala Trp Ser Lys Ile Pro
                  930                                           960
GAG AAA TTG GCT TTT TAT GAT TAC ATT GGA AAT AAT CCG GCG AAA GGG GGA TTA
Glu Lys Leu Ala Phe Tyr Asp Tyr Ile Gly Asn Asn Pro Ala Lys Gly Gly Leu
                        990                                         1020
TTT AGA GCA GGC TCA ATG GAC AAC GGC GAT GGA ATA GCT GTT GGA TGG TTA GGA
Phe Arg Ala Gly Ser Met Asp Asn Gly Asp Gly Ile Ala Val Gly Trp Leu Gly
                              1050                                      1080
CAC CCT ATT TTT AGA GAT AAA GAA GGA CGT GAA CTC TTT GTA CGA CGT ATG CCT
His Pro Ile Phe Arg Asp Lys Glu Gly Arg Glu Leu Phe Val Arg Arg Met Pro
                                    1110
ACT TTT TTT GAA ACC TTT CCG GTC GTT TTG ATA GAT GGG GAC GGA ATT GTT CGA
Thr Phe Phe Glu Thr Phe Pro Val Val Leu Ile Asp Gly Asp Gly Ile Val Arg
      1140                                                      1170
GCT GAT GTT CCT TTT AGA AGA GCA GAA TCT AAG TAT AGC GTT GAA CAA GTA GGT
Ala Asp Val Pro Phe Arg Arg Ala Glu Ser Lys Tyr Ser Val Glu Gln Val Gly
                  1200                                          1230
GTA ACT GTT GAG TTT TAC GGC GGC GAA CTA AAC GGA GTT AGT TAT AGC GAT CCT
Val Thr Val Glu Phe Tyr Gly Gly Glu Leu Asn Gly Val Ser Tyr Ser Asp Pro
                        1260                                        1290
GCT ACT GTC AAA AAA TAT GCT AGA CGT GCT CAA TTA GGT GAA ATT TTC GAA TTA
Ala Thr Val Lys Lys Tyr Ala Arg Arg Ala Gln Leu Gly Glu Ile Phe Glu Leu
                              1320                                      1350
GAC CGT GCT ACT TTG AAA TCA GAT GGT GTT TTT CGC AGT AGT CCG AGG GGT TGG
Asp Arg Ala Thr Leu Lys Ser Asp Gly Val Phe Arg Ser Ser Pro Arg Gly Trp
                                    1380
TTT ACT TTT GGA CAT GCT TCC TTT GCC CTA CTC TTC TTC TTC GGA CAC ATT TGG
Phe Thr Phe Gly His Ala Ser Phe Ala Leu Leu Phe Phe Phe Gly His Ile Trp
      1410                                                      1440
CAT GGG TCT AGA ACC TTG TTC AGA GAT GTT TTT GCT GGG ATT GAC CCC GAT TTG
His Gly Ser Arg Thr Leu Phe Arg Asp Val Phe Ala Gly Ile Asp Pro Asp Leu
                  1470                                          1500
GAT GTT CAA GTA GAA TTT GGC GCA TTC CAA AAA ATT GGA GAT CCA ACT ACA AGA
Asp Val Gln Val Glu Phe Gly Ala Phe Gln Lys Ile Gly Asp Pro Thr Thr Arg
                        1530
AGA CAA GGA GTT TAA TACAACATTGCTTTATGATTTTTTGCCTCTATTTTTTTATGGTTTGACATA
Arg Gln Gly Val -c-
  1580                                                          1630
CGATACTAGAGCAATCTTGATTTAAATCACCGCCTTTTCTTTTTTTTTTTTCATTATTGGGAAAATAGTC
                                    1680
CCAAGTAAACAGGTATGGAAGCTATAATTGTAAACCACAATCGAATCTATGGAAGCATTGGTTTATACAT
                1730                                                1780
TTCTATTAGTCTCTACTTTACCGATTATTTTTTTTGCGATCTTTTTTCGGGAACCTCCTAAAATTTCAAC
                                                    1830
TAAAAAATGAAATGATTTTTCATTATTTCAATTGACGTAATGAGCCTTCCAATATTGGAAGGCTCATTCG
      1860
TCGAC
```

Figure 11: DNA sequence of the non-transcribed strand of the P680 chlorophyll a apoprotein gene and flanking regions from S. oleracea. Corresponding amino acids are indicated.

Other remarkable features of the DNA sequence include (1) a seven base pair repeat TTGTTCC present three times within a span of sixty-six nucleotides between the putative promotor region and the start of the structural gene (Figure 11, positions -167 to -161, -149 to -143, and -108 to -102); this sequence is also present once within the structural gene (position 578 to 584); (2) pair of nine base pair direct repeats of the sequence TATGGAAGC occuring 140 and 164 base pairs distal to the TAA stop codon (Figure 11, positions 1659-1668, and 1693 to 1702), and (3) a stem-loop secondary structure with a seven base pair and a seventeen bp loop located between the putative promoter and the structural gene (Figure 13).

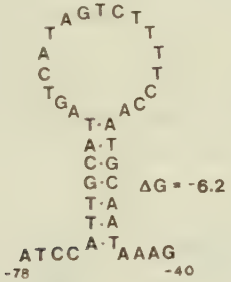


Figure 13: Proposed secondary structure in the 5' flanking sequences to the P680 chlorophyll a apoprotein gene. Numbering is according to Figure 11 and △G was calculated after Tinoco et al (1973).

The codon usage for the 508 amino acids of the P680 chlorophyll a apoprotein gene is tabulated in Figure 14. Although all possible codon species are present for each amino acid, a marked preference (71%) is shown for codons having A or T in the wobble position.

	T	C	A	G	
T	Phe 31	Ser 7	Tyr 10	Cys 2	T
	Phe 10	Ser 4	Tyr 7	Cys 1	C
	Leu 13	Ser 5	Stop 1	Stop -	A
	Leu 13	Ser 5	Stop -	Trp 17	G
C	Leu 4	Pro 13	His 12	Arg 11	T
	Leu 2	Pro 7	His 2	Arg 1	C
	Leu 6	Pro 3	Gln 10	Arg 4	A
	Leu 1	Pro 4	Gln 3	Arg 1	G
A	Ile 17	Thr 13	Asn 5	Ser 10	T
	Ile 3	Thr 5	Asn 3	Ser 3	C
	Ile 10	Thr 6	Lys 11	Arg 11	A
	Met 11	Thr 1	Lys 2	Arg 1	G
G	Val 22	Ala 20	Asp 21	Gly 18	T
	Val 4	Ala 6	Asp 5	Gly 11	C
	Val 11	Ala 10	Glu 20	Gly 23	A
	Val 3	Ala 3	Glu 2	Gly 8	G

Figure 14: Tabulation according to codon specificity of the 508 amino acids of the P680 chlorophyll a apoprotein as deduced from the DNA sequence.

3.6 Amino acid sequence of the P680 chlorophyll a apoprotein

The amino acid composition of the P680 chlorophyll a apoprotein displays the typical highly hydrophobic character (48%) expected for an intrinsic membrane polypeptide. The hydropathy profile calculated according to Kyte and Doolittle (1982) shows domain structure in which the hydropathic residues are clustered into seven discrete core regions probably traversing the membrane in the form of alpha helices (Figure 15). A large hydrophilic domain of nearly 200 amino acid residues, that is more than one-third of the protein, is interspersed between spanning segments VI and VII and contains four pairs of arginine residues.

Spanning segments III and IV, V and VII each contain a pair of carefully spaced histidine residues separated by 13 or 14 amino acids. In addition, a cysteine residue is located between the histidines of segments III and IV (Figure 16).

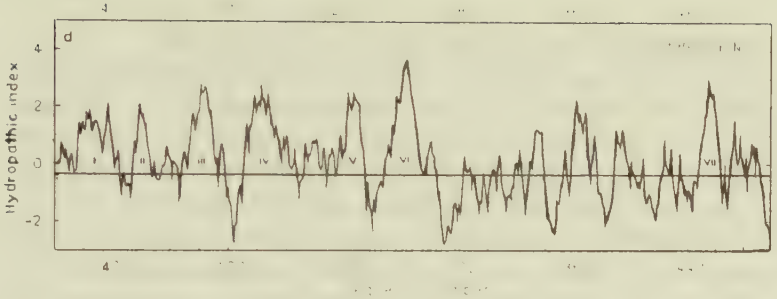


Figure 15: Hydropathy profile of the 508 residue amino acid sequence in which amino acid residue number is plotted against the average hydrophobicity of eleven surrounding residues, according to Kyte and Doolittle (1982). The average of the first eleven putative residues is positioned at the first amino acid. Membrane spanning segments are marked I through VII.

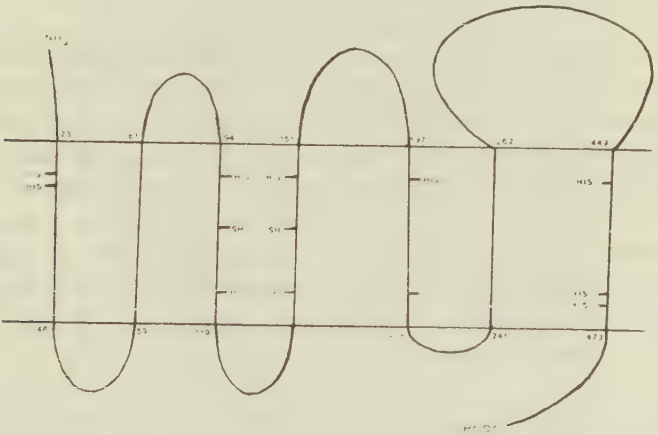


Figure 16: Schema depicting the positioning of the P680 chlorophyll a apoprotein in the thylakoid membrane. Numbers indicate residues where the polypeptide enters or leaves the lipid bilayer. Histidine and cysteine residues which may serve to bind chlorophyll a or pheophytin a are indicated.

3.7 Transcription of the P680 chlorophyll a apoprotein gene

Hybrid selection experiments using immobilized cloned Eco RI subfragments from fragment Sal I-7 and from adjacent regions of fragment Sal I-8 had disclosed that a coherent 7 kbp region was capable of selecting RNA species whose translation products reacted immunologically with antisera to the P680 chlorophyll a apoprotein (Westhoff et al., 1983). Northern blot analysis of spinach plastid RNA using radiolabeled fragments Sal I-7 or Sal I-8 as hybridization partner revealed multiple banded patterns with discrete species ranging from 2.0 to more than 5.0 kb. Hybridization of nitrocellulose strips from the same Northern blots with a large subunit of ribulose bisphosphate carboxylase or 32 kd gene probe gave the known single reacting RNA species of 1.7 (Zurawski et al., 1981) and 1.25 kb (Zurawski et al., 1982), respectively, suggesting that the variable sizes of the RNA species transcribed from Sal I-8/Sal I-7 segments result from posttranscriptional processing or variable read-through of transcriptional termination signals rather than nonspecific secondary degradation.

In order to more exactly define the relationship between these giant RNA species four nick-translated DNA fragments were hybridized to Northern strips of spinach plastid RNA from plant material which was actively synthesizing thylakoid membranes and hence would be expected to possess a high concentration of RNA species for thylakoid membrane proteins (Figure 17). Fragment A contains 160 bp at the amino terminus of the P680 chlorophyll a apoprotein, fragment B 622 bp at the carboxyl terminus plus 340 bp of 3' flanking sequence of the P680 chlorophyll a apoprotein gene, fragment C 600 bp of the cytochrome b_6 structural gene and fragment D 400 bp of the subunit 4 gene.

All four fragments hybridize to three RNA species larger than 4 kb and to a component of 2.5 kb although fragments A and D are more than 4 kb apart. In addition, fragments A and B hybridize with components of 2000 and 2100 bp, fragments C and D with components of 1500 and 1300 bp, and fragment C alone with a species around 900 bp in size (Figure 18).

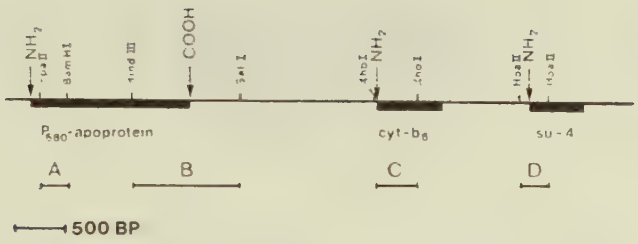


Figure 17: Schema showing the location of fragments A, B, C, and D in relation to the genes for the P680 chlorophyll a apoprotein, cytochrome b_6 and subunit 4 (su-4).

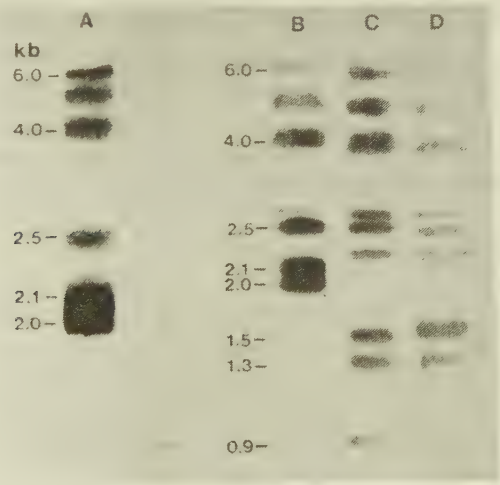


Figure 18: Hybridization between radiolabeled fragments A-D and Northern blots of spinach plastid RNA. A) Fragment Bam HI-Hpa II, 160 bp; B) Fragment Sal I-Hind III, 955 bp; C) Fragment Xho I-17, 295 bp; D) Fragment Hpa II-Hpa II, 300 bp. Fragment A was hybridized to a Northern blot from a different gel.

The 5' ends of P680 chlorophyll a apoprotein transcripts were analyzed using S1 nuclease. The 5' termini of the 335 bp Bam HI fragment or of the 280 bp Taq I fragment (Figure 19) were labeled with γ^{32}P-ATP, hybridized with plastid RNA at 48°C and then treated with S1 nuclease. The resulting fragments which retained a labeled end were detected by autoradiography after separation on 5% PAA gels containing 7M urea.

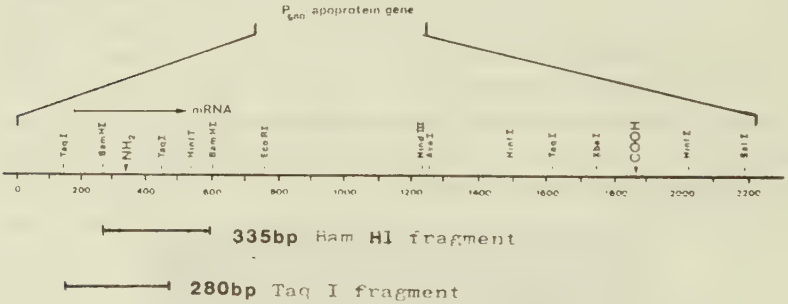


Figure 19: Diagram showing the location of the Bam HI and Taq I fragments used in S1 nuclease experiments in relation to the P680 chlorophyll a apoprotein structural gene.

Figures 20 and 22 show the autoradiograms from the Bam HI and Taq I experiments, respectively. Lane (1) contains the pattern of radioactive DNA fragments after hybridization with RNA and S1 nuclease treatment; lane (2) shows the pattern after hybridization in the absence of RNA but with S1 nuclease treatment; lane (3) shows the untreated DNA sample.

In the Bam HI experiment, lane (1) contains four bands having sizes of 335, 300, 200 and 150 bp. Three of these bands are present in the S1 treated DNA (lane 2). Only the full size fragment is seen in the untreated DNA sample (lane 3).

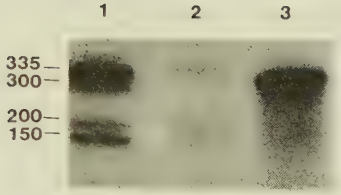


Figure 20: S1 nuclease digestion experiment with Bam HI fragment. Lanes (1) DNA protected after hybridization with RNA and S1 nuclease digestion, (2) DNA protected after hybridization without RNA and S1 nuclease digestion, (3) untreated radiolabeled DNA.

The 35 bp missing from the second protected fragment probable result from an S1 nuclease cutting activity just at the base of a hairpin loop structure (Figures 13 and 21). This structure appears to be a RNA-DNA duplex since it is not present in lane (2); the two smaller fragments probably result from S1 cutting activity in two DNA-DNA duplex structures formed in GC rich regions within the Bam HI fragment. The three S1 cutting sites are depicted in Figure 21.

Figure 21: Three possible S1 nuclease cutting sites within the 335 bp Bam HI fragment. Numbering follows Figure 11.

Since the 335 bp Bam HI fragment also appears to be protected in its full length, the 5' end of at least one P680 transcript must lie in sequences 5' to the leftmost Bam HI site.

In Figure 22, the autoradiogram from the corresponding experiment with the Taq I fragment is shown. Two bands are present in lane (1) corresponding to the fully protected fragment (280 bp) and an S1 nuclease cutting activity at the base of the stem-loop structure, as seen with the Bam HI fragment.

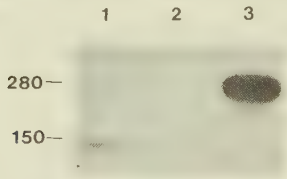


Figure 22: S1 nuclease digestion experiment with Taq I fragment. Lanes (1) DNA protect[illegible]fter hybridization with RNA and S1 nuclease digestion, (2) DNA protected after hybridization without RNA and S1 nuclease digestion, (3) untreated radiolabeled DNA.

It is not clear from these experiments alone whether the S1 cutting activity at the base of the stem-loop structure (Figure 21) creates an artifactual 5' end or whether this 5' end already exists in the population of P680 chlorophyll a apoprotein gene transcripts. However, the full protection of both fragments indicates the existence of transcripts which begin 5' to the stem-loop structure.

4.0 Location and sequencing stradegy for the 44 kd chlorophyll a apoprotein gene from PS II

The gene for the 44 kd chlorophyll a apoprotein has been localized on the spinach chromosome (Figure 4) to fragments Pst I-4, Bam HI-8b, and Bam HI-16 (Westhoff et al., 1983). In preparation for DNA sequencing 50 to 100 µg of each of four fragments were isolated from the cloned material (Herrmann et al., 1983): the leftmost Pst I-Pvu II fragment (1400 bp) from Pst I-4, the leftmost Bam HI-Pvu II fragment (1100 bp) from Bam HI-8b, a Sac I-Sac I fragment (1300 bp) from the left side of Bam HI-8b and the Bam HI-16 fragment (2.3 kb) as indicated in Figure 23.

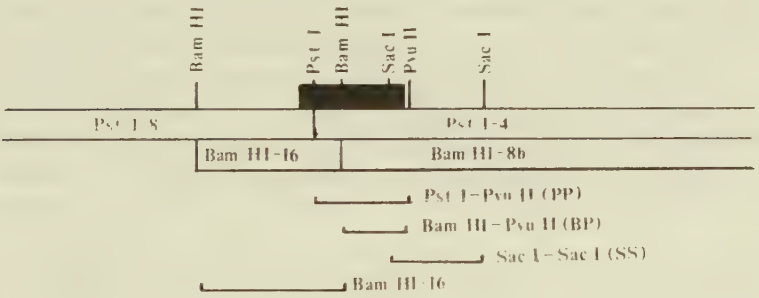


Figure 23: Diagram showing locations of the four fragments isolated from cloned material which contain the 44 kd chlorophyll a apoprotein gene of PS II (heavy line).

Subfragments were isolated from these fragments as indicated in Figure 24: the Hpa II and Hinf I subfragments from the Bam HI-Pvu II fragment (BP); the Hind III and Dde I fragments from the Pst I-Pvu II fragment (PP); the Hpa II, Hinf I, and Sau 3A fragments from the Sac I-Sac I fragment (SS) and the Dde I fragments from Bam HI-16. Fragments with Hpa II ends were labeled with α^{32}P-dCTP, those with Hinf I, Sau 3A, and Bam HI ends with unlabeled dGTP and α^{32}P-dATP; Hind III ends were labeled with α^{32}P-dATP and Dde I ends were labeled

with unlabeled dTTP and dCTP and α^{32}P-dATP. Labeled fragments were tested with a battery of enzymes (Acc I, Alu I, Dde I, Hae III, Hinf I, Hha I, Hind III, Pst I, Pvu II, Sac I, Sau 3A, and Taq I) to locate a suitable second cut site (when both ends were labeled) and to establish a restriction map. The exact ordering of the labeled fragments was determined by DNA sequence overlap.

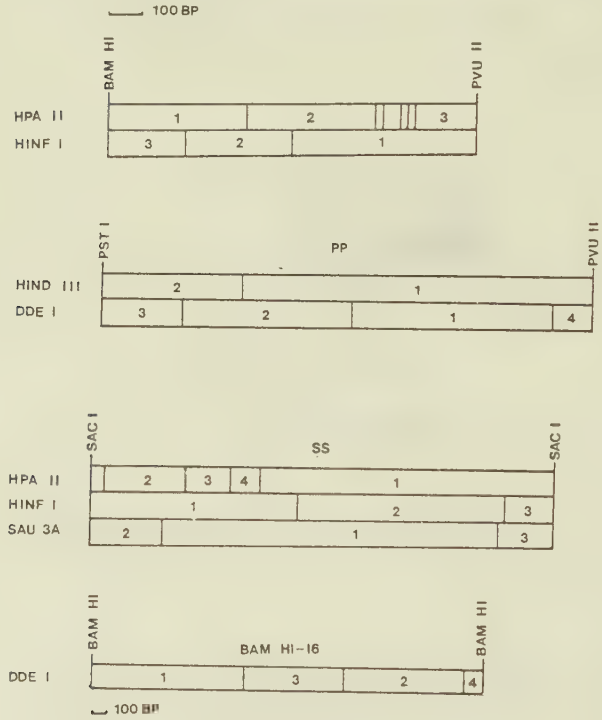


Figure 24: Subfragments used for DNA sequencing isolated from the four fragments in Figure 23. Numbering is in order of decreasing size. Fragments smaller than 60 bp are not numbered. The order of Dde I fragments 1 and 3 in Bam HI-16 is not certain.

Figure 25 shows a partial restriction map of the region of the spinach plastid chromosome encoding the 44 kd chlorophyll a apoprotein gene. The direction and extent of the individual sequence analyses are indicated by the arrows and the numbered fragments are listed in Table 3.

4.1 Nucleotide sequence of the 44 kd chlorophyll a apoprotein gene

In Figure 26 the nucleotide sequence of a continuous interval of 2800 bp including the structural gene for the 44 kd chlorophyll a apoprotein and flanking regions and about half of the D2 protein gene is presented. The sequence is translated in all six possible reading frames. Analysis of the nucleotide sequence reveals that there is one large open reading frame of 473 triplets in the expected position and orientation for the 44 kd chlorophyll a apoprotein gene. This reading frame starts within fragment Bam HI-8b, 984 nucleotides upstream of the Bam HI site separating Bam HI-8b and Bam HI-16, and ends at an amber codon at nucleotide 2429 just after the unique Pst I cleavage site within fragment Bam HI-16. All other reading frames within this DNA sequence include multiple termination codons interspersed throughout the sequence.

Table 4 summarizes the codon usage of the 473 triplets encoding the 44 kd chlorophyll a apoprotein.

	T		C		A		G	
T	Phe	21	Ser	12	Tyr	12	Cys	4	T
	Phe	12	Ser	2	Tyr	3	Cys	0	C
	Leu	20	Ser	1	Stop	0	Stop	1	A
	Leu	11	Ser	5	Stop	0	Trp	17	G
C	Leu	14	Pro	12	His	10	Arg	6	T
	Leu	1	Pro	9	His	3	Arg	4	C
	Leu	9	Pro	6	Gln	8	Arg	2	A
	Leu	2	Pro	2	Gln	0	Arg	3	G
A	Ile	13	Thr	10	Asn	10	Ser	7	T
	Ile	5	Thr	8	Asn	4	Ser	0	C
	Ile	8	Thr	2	Lys	10	Arg	5	A
	Met	8	Thr	1	Lys	3	Arg	4	G
G	Val	6	Ala	23	Asp	9	Gly	18	T
	Val	7	Ala	6	Asp	5	Gly	10	C
	Val	13	Ala	10	Glu	18	Gly	19	A
	Val	6	Ala	1	Glu	4	Gly	12	G

Table 4: Codon usage in the 44 kd chlorophyll a apoprotein gene

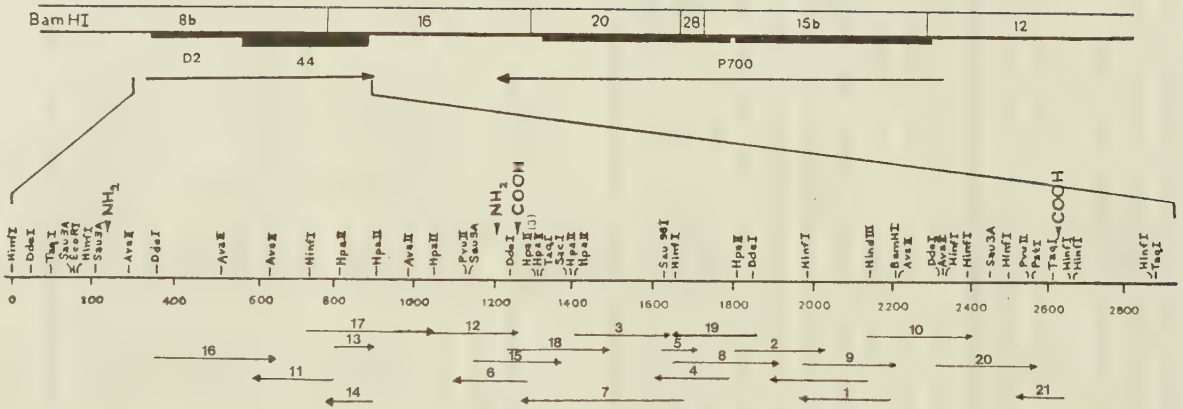


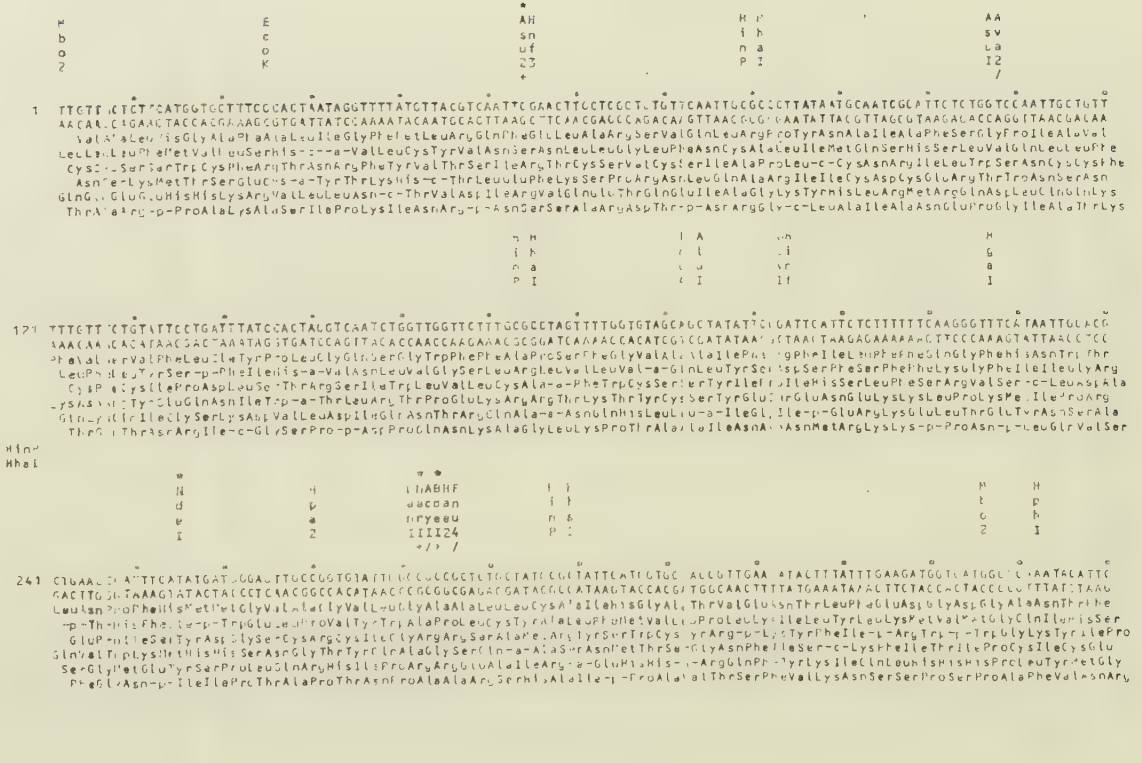
Figure 25: Partial restriction map and sequencing stradegy diagram for the 44 kd chlorophyll a apoprotein gene of PS II on the spinach plastid chromosome. Positions of the D2, 44 kd chlorophyll a apoprotein and two P700 chlorophyll a apoproteins are indicated by thickened lines; their direction of transcription is indicated by thickened arrows. Small arrows indicate the direction and extent of individual sequence analyses; numbers of fragments refer to Table 3.

TABLE 3: DNA FRAGMENTS FOR SEQUENCE ANALYSIS OF THE 44 KD CHLOROPHYLL A APOPROTEIN GENE OF PS II

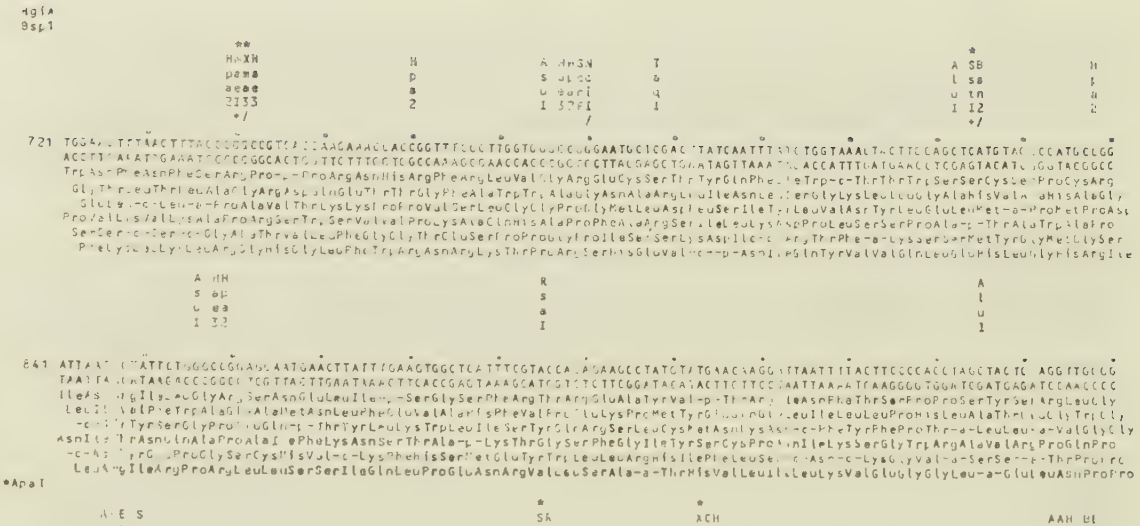
NO.	FRAGMENT	SIZE (BP)	LABELED END/SECOND CUT
1	Bam HI-8B	4400	Bam HI/Sac I
2	BP/Hpa II-1	400	Hpa II/Hinf I
3	BP/Hpa II-2	390	Hpa II/Hinf I
4	BP/Hpa II-2	390	Hpa II/Hinf I
5	BP/Hpa II-2	390	Hpa II/Hinf I
6	BP/Hpa II-3	170	Hpa II
7	BP/Hinf I-1	550	Hinf I
8	BP/Hinf I-2	350	Hinf I/Acc I
9	BP/Hinf I-3	240	Hinf I
10	PP/Hind III-2	400	Hind III
11	SS/Hpa II-1	850	Hpa II
12	SS/Hpa II-2	250	Hpa II/Dde I
13	SS/Hpa II-3	140	Hpa II/SS*
14	SS/Hpa II-3	140	Hpa II/SS*
15	SS/Sau 3A-2	220	Sau 3A
16	SS/Sau 3A-1	900	Sau 3A/Taq I
17	SS/Hinf I-1	610	Hinf I/Pvu II
18	PP/Dde I-1	600	Dde I/Hinf I
19	PP/Dde I-1	600	Dde I/Hinf I
20	Bam HI 16/Dde I-2	700	Dde I/Pst I
21	Bam HI 16/Hinf I	200	Hinf I/Pvu II

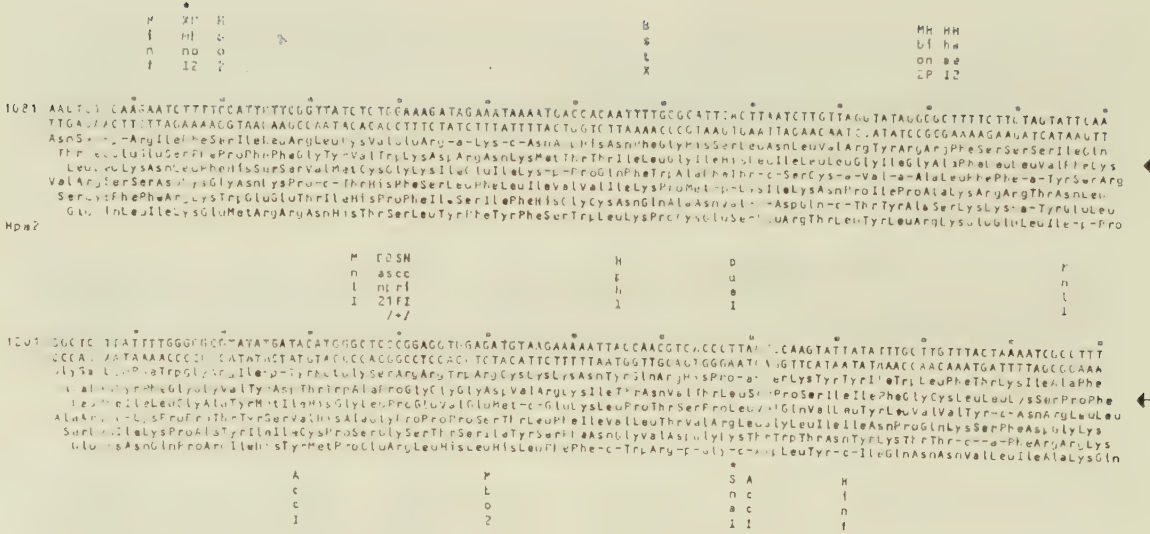
BP, PP, SS, BAM HI 16 = Abbreviations for fragments as in Figure 24.

* SS= Strand Separation.



75




HpaII

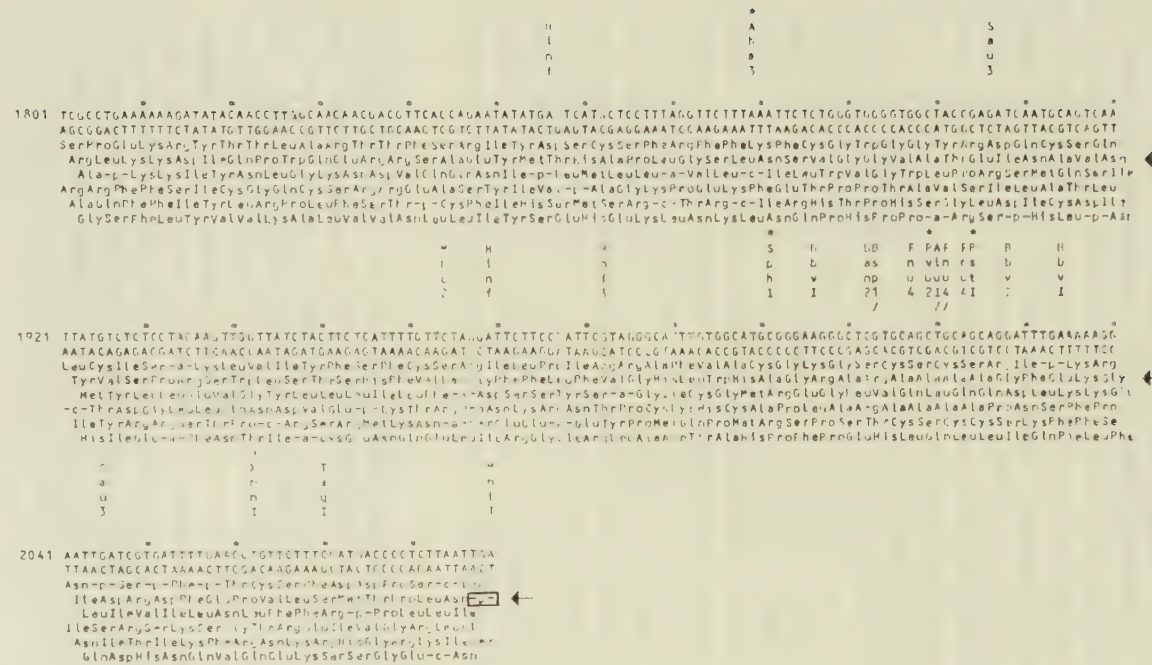


Figure 26: Nucleotide sequence of 2090 bp of spinach plastid DNA including the gene for the 44 kd chlorophyll a apoprotein of the PS II reactior center. Positions of selected restriction sites are shown and the sequence is translated in all six possible reading frames. The position of the open reading frame coding for the 44 kd chlorophyll a apoprotein gene is marked with arrows.

4.2 Amino acid sequence of the 44 kd chlorophyll a apoprotein

In Figure 27 the deduced amino acid sequence for the 44 kd chlorophyll a binding apoprotein of PS II is presented. The site of translational initiation for the 44 kd chlorophyll a apoprotein is ambiguous and cannot be determined from DNA sequence data alone. The deduced amino acid sequence, beginning at the first possible methionine residue (ATG) in the open-reading frame would form a polypeptide of 51.8 kd which is about 6 kd larger than the predicted molecular weight of the authentic protein and full-chain products obtained by cell-free translation of spinach mRNA or transcription-translation of plastid DNA restriction fragments. If, on the other hand, translation were to commence at the second methionine the calculated molecular mass of 44.6 kd would be closer to the reported 43 or 44 kd of the authentic protein. Possible ribosome binding sites can be identified 5' to each of these methionine codons. The 44 kd chlorophyll a apoprotein does not appear to be made as a precursor that undergoes further proteolytic cleavage (Alt et al., 1983). It is probable that translation begins at the first methionine codon since molecular weights of hydrophobic proteins estimated by SDS PAA gel electrophoresis are known to be inaccurate (Morris and Herrmann, 1984; Kirsch et al., in prep.; Zurawski et al., 1982; Hearst and Sauer, 1984). Like other known chloroplast encoded protein genes the 44 kd chlorophyll a apoprotein gene shows a striking preference for A and T (70%) in the third position of the degenerative codons.

About 50% of the amino acids of the 44 kd chlorophyll a apoprotein are hydrophobic. The deduced amino acid sequences were therefore analysed by computer according to the procedure of Kyte and Doolittle (1982) for local hydrophobicity and the distribution of amino acid residues that may be important in pigment binding and protein folding (Figure 28). As with other membrane bound proteins, the hydrophobic residues tend to be clustered in stretches of 20 to 25 residues which probably represent transmembrane spans. These spans are connected by hydrophilic turn segments that appear to be alternatively exposed to stroma and thylakoid lumen. Single or paired charged residues frequently flank these spans and probably serve to stabilize the posi-

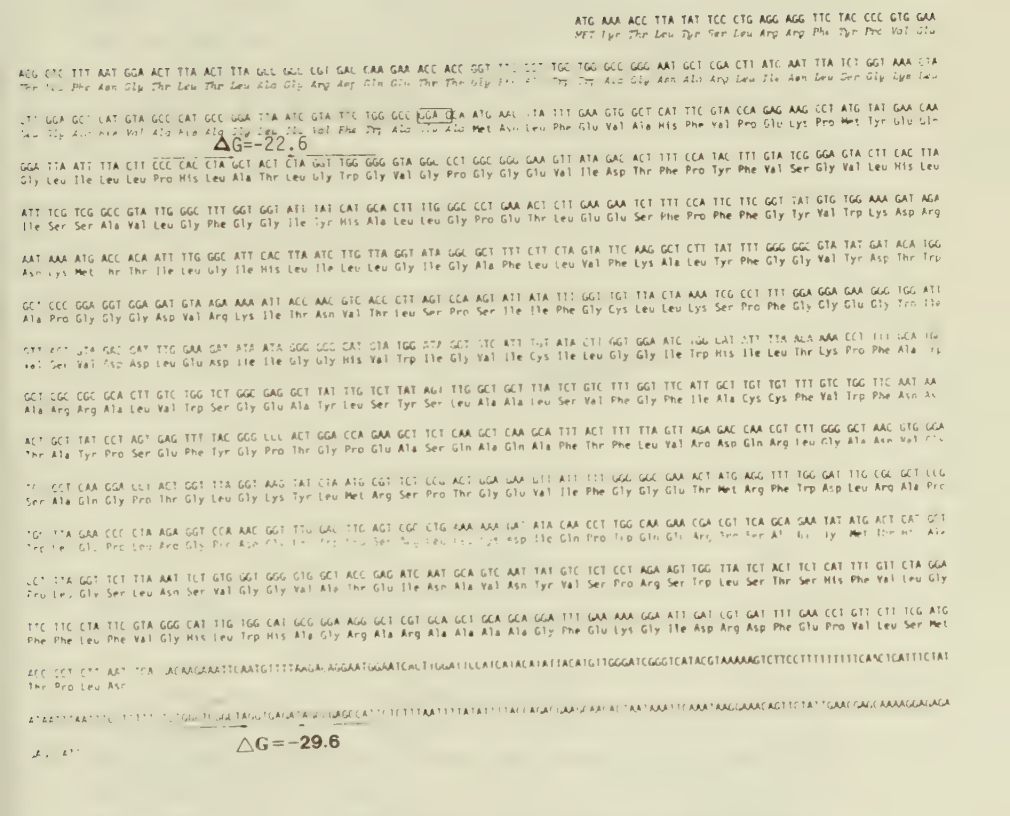


Figure 27: Nucleotide sequence of the interval of the spinach plastid chromosome encoding the 44 kd chlorophyll a apoprotein of the PS II reaction center. The non-coding strand is presented; protein-coding parts are written in triplets, possible translational initaition sites for for the protein are discriminated by italic letters. Regions homologous to the 3'end of maize and tobacco plastid 16S rRNA are boxed. Elements capable of forming stable hairpin structures are marked by arrows.

tioning of the protein chain in the lipid matrix. The 44 kd chlorophyll a apoprotein has seven hydrophobic spans implying that the amino and carboxyl termini are on opposite sides of the membrane.

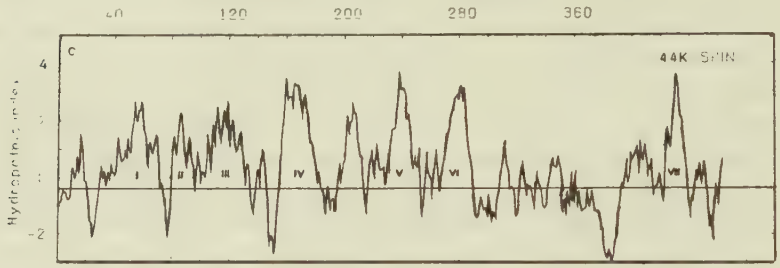


Figure 28: Hydropathy profile for the 44 kd chlorophyll a apoprotein.

In the 44 kd chlorophyll a apoprotein there are three pairs of histidine residues located in hydrophobic stretches; in each case the His residues are separated by 13 amino acid residues (Figure 29).

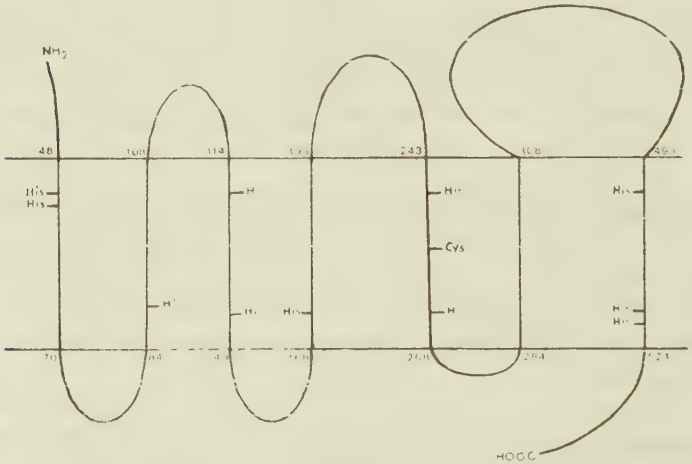


Figure 29: Schematic representation of the 44 kd chlorophyll a apoprotein positioned in the thylakoid membrane.

5.0 The 32 kd herbicide-binding gene from Amaranthus

The gene for the 32 kd herbicide-binding protein from spinach has been localized to Sal I fragment 6 (Figure 4, Driesel et al., 1980) and sequenced (Zurawski et al., 1982). In addition, the Eco RI-Sal I subfragment containing the gene has been subcloned into pBR322 (Schmidt, unpublished results). A partial restriction map could be constructed from the known pBR322 sequences and those of the gene:

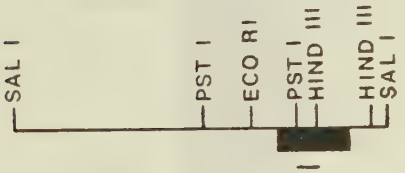


Figure 30: Partial restriction map of the clone pWHsp 206/ES containing the 32 kd gene from spinach (thickened black line).

A Pst I-Hind III double digest gives four fragments having sizes of 3100, 1500, 880, and 240 bp. The latter fragment (underlined in Figure 30) contains sequences only internal to the gene and was used to detect the fragment containing the gene in a Sal I pattern of Amaranthus retroflexus plastid DNA. Amaranthus plastid DNA was digested with Sal I, the frag-ments separated on a 1% agarose gel, transferred to nitrocellulose and hybridized to the ^{32}P.labeled 240 bp fragment. Figure 31 shows the Sal I restriction fragment pattern and the accompanying autoradiogram. The 240 bp spinach probe hybridized to a 9.6 kb fragment as seen in the accompanying autoradiogram.

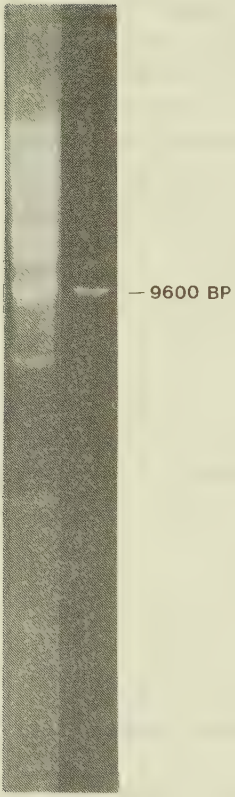


Figure 31: Sal I restriction pattern of plastid DNA from Amaranthus retroflexus separated on a 1% agarose gel. The 240 bp spinach probe hybridizes to a 9.6 kb fragment as seen in the negative image reproduced from the autoradiogram.

5.1 Cloning of the 32 kd gene from a herbicide-resistant strain of Amaranthus retroflexus

The cloning of the plastid DNA from triazine-resistant plants of Amaranthus retroflexus is described in Section 2.6.1. The transformation procedure yielded approximately 20,000 transformants from 10^7 E. coli cells. Of 936 transformants which were checked, 699 (75%) showed chloramphenicol sensitivity indicating that they carry an Eco RI insert. Although the transformation rate was not particularly high, a high percentage of the transformants contain hybrid plasmids indicating the effectiveness of the vector-phosphatase treatment method.

Further screening of the 699 clones by the method of Grunstein-Hogness using the 240 bp spinach fragment as hybridization partner (Section 2.6.1) yielded six clones containing an Eco RI insert with the 32 kd gene. The recombinant strain pMHamE was purified as a single colony, and its plasmid DNA isolated by alkali lysis from six liters of culture medium.

5.2 Restriction map of pMHamE

A restriction map could be constructed after single and double digestions of the plasmid DNA with Hind III, Pst I, and Xba I. Figure 32 shows the pattern of fragments seen after various digestions and separation on an agarose gel. The accompanying schema shows the order of fragments.

5.3 DNA sequencing of the 32 kd gene

A Hind III digest of pMHamE produces four fragments of 6600, 3700, 880, and 550 bp (Figure 31). The two smallest fragments are contingent; the Hind III site joining them is within the 32 kd gene. The two fragments were labeled with α^{32}P-dATP. To separate the two labeled ends, the 550 bp fragment was cut with Pst I or Hinf I giving subfragments of 240 and 310 bp or 30 and 520 bp, respectively. The 880 bp fragment was cut with Taq I giving labeled subfragments of 240 and 420 bp. The subfragments were separated on 5% PAA, eluted and sequenced.

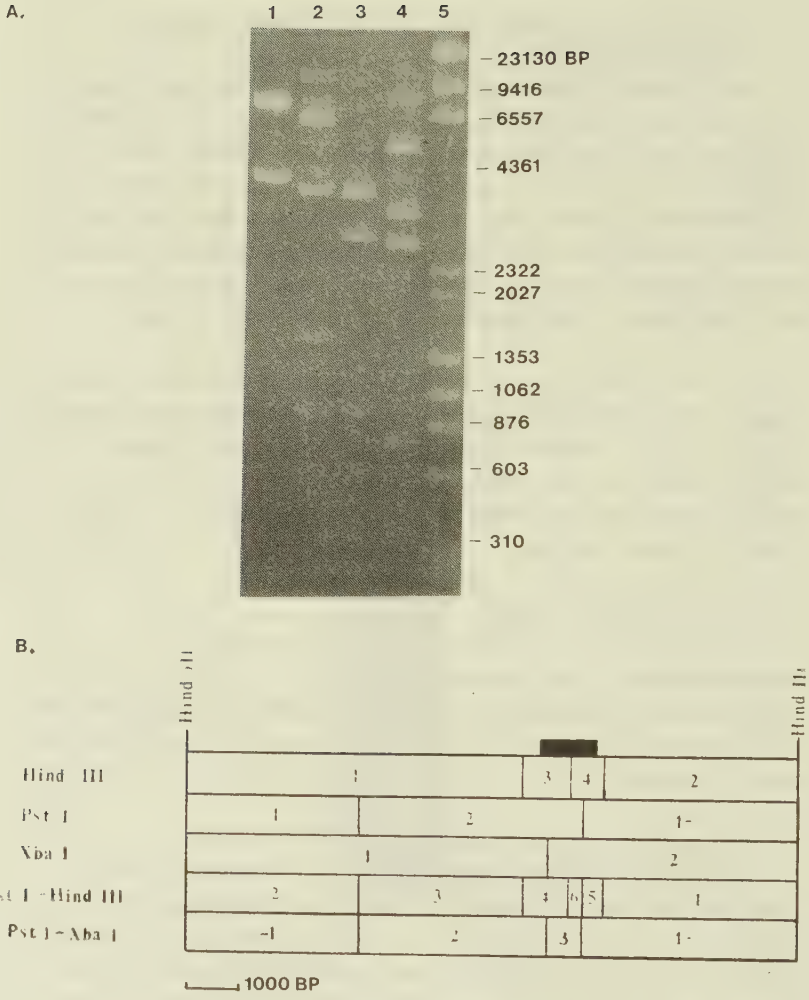


Figure 32: (A) Photograph of a 0.8% agarose gel in which fragments from (2) Hind III. (3) Pst I-Hind III. (4) Pst I-Xba I, (5) Hind III digest of lambda DNA and Hinf I digest of pBR322; (B) Diagram showing the order of the fragments for the pMMamE plasmid. Position of gene is marked with a thick black line.

An XbaI-Pst I double digest of pMHamE yields three fragments having lengths of 7700, 3500 and 760 bp. The latter fragment is a Pst I-Xba I fragment in which both ends are located within the gene. It was selectively labeled at the Xba I site with α^{32}P-dCTP and directly sequenced. This sequence was run in parallel with that from the corresponding fragment from the wild-type spinach gene. Among the nucleotide differences seen between these two genes only one leads to an amino acid change: a serine to glycine change at amino acid 264. The individual sequence analyses are depicted schematically in Figure 33. In Figure 34 the DNA sequence beginning fifty base pairs before the met codon at the amino terminus and ending with a TAA stop codon at the end of the structural gene is given. Figure 35 shows a photograph of the sequencing gel comparing the wildtype spinach sequence with the corresponding sequence from the herbicide-resistant strain of Amaranthus. The nucleotide change responsible for the amino acid change is indicated. Figure 36 shows the hydropathy profile calculated according to Kyte and Doolittle (1982) with an interval of 11 amino acids. For comparison, the corresponding hydropathy curve for the D2 protein is shown (Alt et al., 1984).

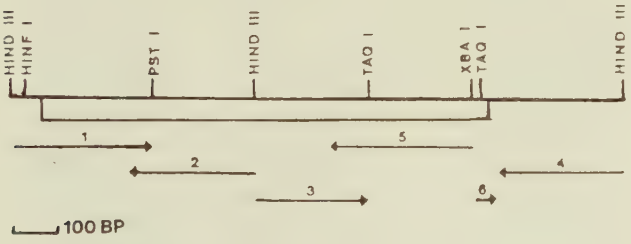
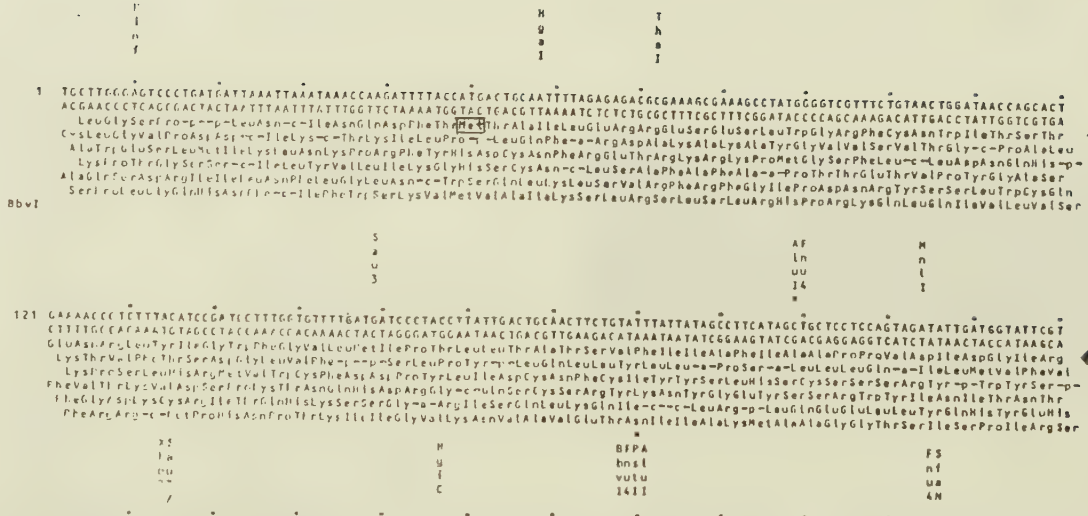


Figure 33: Sequencing strategy for the 32 kd gene from Amaranthus retroflexus.


BbvI
121
241

721
DdvI
841
DdvI
961

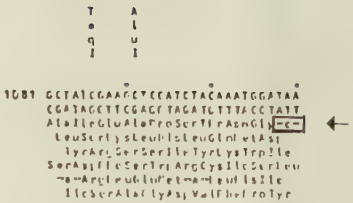


Figure **34**: Nucleotide sequence of 1110 bp of DNA from the plastid chromosome of a herbicide-resistant species of Amaranthus retroflexus. The sequence begins 50 bp in front of the structural gene for the 32 kd protein of PS II and ends at the TAA stop codon for this gene. The open-reading frame is marked with arrows. Positions of selected restriction sites are shown and the sequence is translated in all possible reading frames.

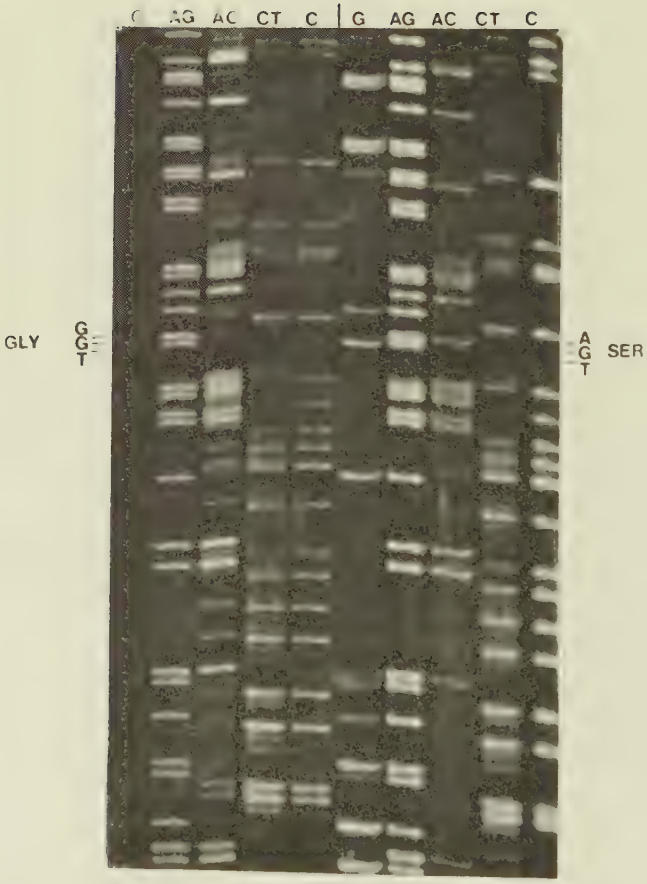


Figure 35: Photographic reproduction of the sequencing gel in which the wild-type spinach 32 kd gene is compared with the corresponding gene from a herbicide-resistant strain of Amaranthus retroflexus. The five base-specific cleavage reactions are indicated. The A. retroflexus sequence on the left codes for a gly residue (GGT) while the spinach sequence on the right, like the wildtype Amaranthus hybridus 32 kd gene, codes for a ser residue (AGT) at this position.

5.4 Amino acid sequence of the 32 kd protein from herbicide-resistant strain of Amarnathus retroflexus

The 32 kd gene from the herbicide-resistant strain of Amaranthus retroflexus contains an open reading frame capable of coding for a protein having 373 amino acid residues starting with the first methionine residue. Figure 37 shows the deduced amino acid sequence of the 32 kd herbicide binding protein from the Amaranthus retroflexus triazine resistant strain. The primary sequence of the protein shows complete conservation with that of spinach, tobacco (Zurawski et al., 1982), maize (McIntosh and Bogorad, unpublished), soybean (Spielmann and Stutz, 1983) and S. nigrum (Hirschberg et al., 1984) except for a single amino acid change from ser to gly at position 264 (underlined). This residue appears to be located just inside the lipid bilayer in the seventh membrane spanning section of the protein. The wildtype 32 kd protein of Amaranthus hybridus has one neutral amino acid substitution, e.g. ile for val, at position 281.

5.5 The tRNA HIS gene

The 420 bp subfragment obtained after cutting the 920 bp Hind III fragment with Taq I contains the $tRNA^{HIS}$ gene depicted in Figure 38. The histidine tRNA gene of Amaranthus retroflexus is located in the same relative position as the maize chloroplast $tRNA^{HIS}$ and has the same anticodon GTG at positions 36-38 (Scwarz et al., 1981). The ribonucleotide sequence deduced from the chloroplast $tDNA^{HIS}$ (Figure 38) shows about 60% homology with the $tRNA^{HIS}$ of E. coli (Singer and Smith, 1972) and about 48% homology to that of yeast mitochondria (Gauss and Sprinzl, 1981; Bos et al., 1979). All highly conserved positions (Gauss and Spinzl, 1981) are maintained in the chloroplast $tDNA^{HIS}$ sequence. The major differences between the $tRNA^{HIS}$ of E. coli and that of the chloroplast occur in the D loop and stem. The sequence of the extra arm is also different from that of E. coli but its size is the same. Unlike genes for tRNAs of E. coli, the universal 3'-terminal tRNA triplet CCA is not a part of the chloroplast gene. This sequence is added posttranscriptionally in coliphages, eukaryotic cytoplasmic tRNAs, yeast mitochondrial tRNAs and chloroplast tRNAs (Whitfeld and Bottomley, 1983).

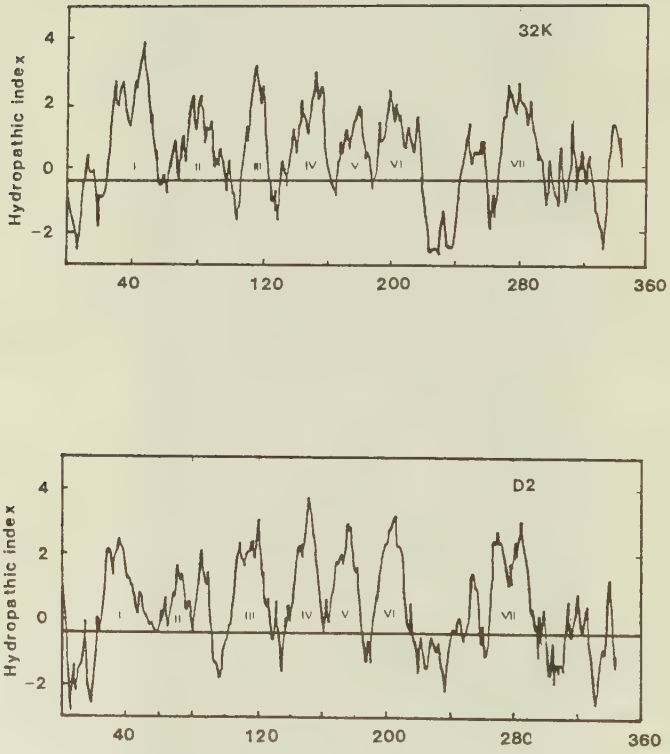


Figure 36: (a) Hydropathy profile for the herbicide resistant form of the 32 kd protein from Amaranthus retroflexus, (b) for comparison, the same hydropathy profile for the D2 protein from spinach (Alt et al., 1984). An interval of elevan amino acids was used. Roman numbers represent regions sufficiently large to span the membrane lipid bilayer.

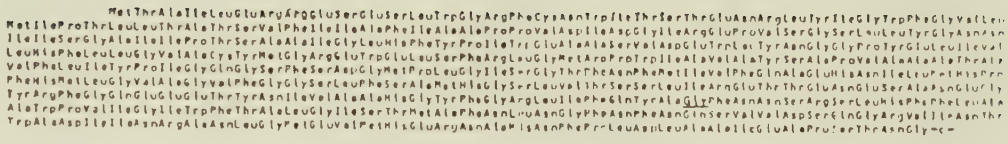


Figure 37: Amino acid sequence of the 32 kd protein from a herbicide-resistant strain of Amaranthus retroflexus. The single amino acid difference between this strain and the wildtype sequence of the corresponding protein of Amaranthus hybridus (Hirschberg and McIntosh, 1983) is a Ser to Gly (underlined) at position 264 caused by a single nucleotide change from A to G.

(A)

5'-TTGCAAAATGATATAACACTTATGAACGAATTTTTAATCAATGAATCGAAGGCTTAACAAAG

GCAAGTTCTGAATACTGAACTTACAAATAAATTCCGAATTTTGGATAGGGGGCGGATGTA

GCCAAGTGGATCAAGGCAGTGGATTGTGAATCCACCATCGCGCGGGTTCAATTCCCGTCGTT

CGCCCATCCAAACCATCCAAATTCAACAT-3'

(B)

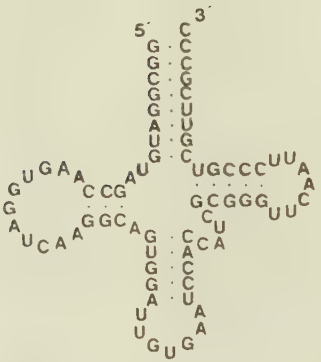


Figure 38: (A) Nucleotide sequence of the section of the Amaranthus retroflexus plastid chromosome encoding the $tRNA^{HIS}$ gene (underlined); (B) Deduced ribonucleotide sequence of the $tRNA^{HIS}$ folded in the stem loop configuration.

6.0 Discussion

The thylakoid membrane of higher plants is composed of approximately 45 unique polypeptide species. Most of the proteins are associated with six supramolecular complexes embedded in the membrane which participate in photosynthetic electron transport (Figure 2). Two of the complexes are associated with chlorophyll *a* molecules capable of performing the primary photosynthetic charge separation reactions. Approximately half of the genes for the proteins are encoded by the plastid chromosome (Figure 3) including all of the reaction center apoproteins (Westhoff et al., 1983a; Westhoff et al., 1983b; Morris and Herrmann, 1984; Alt et al., 1984) cytochromes (Alt et al., 1983; Heinemeyer, 1983; Alt and Herrmann, 1984) and five ATP synthase subunits (Westhoff et al., 1981; Krebbers et al., 1982; Howe et al., 1982; Alt et al., 1983; Herrmann et al., 1983).

The biogenesis of the thylakoid membrane is under environmentally responsive genetic control and develops rapidly in the light (Herrmann and Possingham, 1980). Since the membrane is composed of proteins coded in separate genetic compartments of the plant cell, mechanisms must exist for the coordinated expression of photosynthetic genes such that delivery of the thylakoid components in near stoichiometric amounts can be achieved for efficient assembly of the membrane complexes. Although the organization of plastid genes has been studied extensively during the past several years, relatively little is known about the mechanisms of chloroplast gene transcription or the control of gene expression at different stages of development.

All plastome-coded thylakoid membrane proteins are products of single genes scattered throughout the 70 kb single-copy region of the circular chromosome (Figure 3). The genes are found on both DNA strands and do not appear to be interrupted by introns (Herrmann et al., 1985). In spinach and other higher plants, chloroplast genes coding for proteins which belong to the same thylakoid membrane complex do not appear to be organized in operons. The CF_0-III gene (Alt et al., 1983) and the CF_1 alpha subunit gene (Westhoff et al., 1981) are some 40 kb away from the CF_1 beta-epsilon subunit genes (Westhoff et al., 1981). Similarly,

although the gene for the reaction center apoprotein of PS II (P680 chlorophyll *a* apoprotein gene; Westhoff et al., 1983) is located close to the cytochrome b_6 and subunit 4 genes (Alt et al., 1983), some 70 kb separate it from the gene for the 32 kd herbicide-binding protein (Zurawski et al., 1982) and the genes for the 44 kd chlorophyll *a* binding apoprotein and the D2 protein of the same membrane complex (Alt et al., 1984) are situated halfway between these two. The cytochrome *f* gene (Alt and Herrmann, 1984) is 15 kd away from the cytochrome b_6 and subunit 4 genes (Alt et al., 1983).

The successful localization and cloning of the genes for thylakoid membrane proteins has paved the way for a more penetrating analysis of both gene and protein fine-structure using the technique of DNA sequencing (Maxam and Gilbert, 1977; 1980). The primary amino acid sequence of the protein which can be deduced from the nucleotide sequence of the structural gene yields information about structure-function relationships important in the process of photosynthesis. In addition, the nucleotide sequences of the 5' and 3' flanking regions of the genes yield information about the control of their expression. Finally, a comparison of primary protein sequence data allows conclusions to be drawn concerning the evolution of photosynthetic membranes. At the beginning of this work, aimed at investigating the two chlorophyll *a* conjugated apoproteins of PS II, only two plastome-coded genes for photosynthetic polypeptides had been sequenced, those for the large subunit of ribulose 1,5 bisphosphate (Zurawski et al., 1981) and the 32 kd herbicide-binding protein (Zurawski et al., 1982) proteins. During the course of this analysis, ATP synthase subunit genes (Howe et al., 1982; Alt et al., 1983; Zurawski et al., 1982; Krebbers et al., 1982; Shinozaki et al., 1983; Deno et al., 1983) and at the end, cytochrome *f* (Willey et al., 1984; Alt and Herrmann, 1984; Willey et al., 1984) and cytochrome *b* genes (Heinemeyer, 1983; Phillips and Gray, 1984), but none of the photosynthetic apoproteins which play a fundamental role in the primary photosynthetic light reactions, were sequenced.

The nucleotide sequence of the gene for the large subunit of ribulose 1,5 bisphosphate has been determined in maize (McIntosh et al.,

1980), in spinach (Zurawski et al., 1981), in tobacco (Shinozaki and Suguira, 1982), in the cyanobacterium Anacystis nidulans (Shinozaki et al., 1983) in the cyanobacterium Anabaena (Curtis and Haselkorn, 1983) and in the green alga Chlamydomonas (Dron et al., 1982). The amino acid sequences among the three angiosperms show 90% homology although the corresponding DNA sequences differ by about 25% with most changes occurring in the third position of the codons and in the 3' terminal portion of the structural gene. The gene for the CF_0-III subunit (DCCD binding proteoplipid) has been sequenced in spinach (Alt et al., 1983) and wheat (Howe et al., 1982). The two amino acid sequences show 100% homology with one another, and still about 35% homology with the comparable bacterial (Gay and Walker, 1981) and mitochondrial (Hensgens et al., 1979) subunits. The nucleotide sequences of the wheat and spinach structural genes differ by 18% (Alt et al., 1983). The genes for the beta and epsilon subunits of the ATP synthase have been sequenced in spinach (Zurawski et al., 1982), maize (Krebbers et al., 1982) and tobacco (Shinozaki et al., 1983). The amino acid homology between the three beta subunits ranges between 91 and 96% and for the epsilon subunits between 70 and 80%. The homology between beta and epsilon subunits from tobacco with those of E. coli is still 62% and 27% respectively (Shinozaki et al., 1982). The beta subunit termination codon overlaps the epsilon subunit initiation codon in the three species above but not in pea (Whitfeld et al., 1983). The nucleotide sequence of the gene for the alpha subunit of the ATP synthase has been determined in tobacco (Deno et al., 1983). The deduced amino acid sequence shows 54% homology with its E. coli counterpart. From the cytochrome b_6/f complex, the nucleotide sequence of the cytochrome f gene has been determined in wheat (Willey et al., 1984), in pea (Willey et al., 1984) and in spinach (Alt and Herrmann, 1984). The nucleotide sequences of the wheat and pea genes show 85% homology and the corresponding amino acid sequences 88% homology with the heme binding site and the hydrophobic C-terminal sequence being most highly conserved (Willey et al., 1984). The genes for cytochrome b_6 and subunit 4 polypeptides, which show homologies to the mitochondrial cytochrome b subunit around the heme binding sites (Widger et al., 1983) have been sequenced in spinach (Heinemeyer, 1983). A 15.2 kd polypeptide gene which probably corresponds to subunit 4 has

been sequenced in pe (Phillips and Gray, 1984). The gene for the herbicide-binding 32 kd membrane protein has been sequenced in spinach and tobacco (Zurawski et al., 1982), soybean (Spielmann and Stutz, 1983), a wildtype and herbicide resistant strain of Amaranthus hybridus Hirschberg and McIntosh, 1983) and mustard (Link and Langridge, 1984). All six genes have an open reading frame capable of coding for a protein with 353 amino acids (M approximately 39,000) with only one divergent residue.

The nucleotide sequence of the P680 chlorophyll a apoprotein gene (Figure 11) and the 44 kd chlorophyll a apoprotein gene (Figure 27) of spinach plastid DNA reveals large non-interrupted open reading frames of 1524 and 1419 bp, respectively. Their deduced amino acid sequences code for a 508 and a 473 residue protein of M_r 56,246 and 51,785, respectively. On the basis of similarities in protein sequence and secondary structure (see below) one can infer that these open reading frames code for the 51 and 44 kd apoproteins. These figures deviate significantly from the apparent molecular weights of 51 kd and 44 kd for the chlorophyll a-conjugated proteins associated with PS II estimated according to electrophoretic mobility in SDS-containing PAA gels (Westhoff et al., 1983). The proteins do not appear to be synthesized as precursors after in vitro translation using either an RNA-programmed rabbit reticulocyte system or a DNA-programmed E. coli assay (Westhoff et al., 1983); the largest components reacting immunologically with antisera elicited against the 51 and 44 kd proteins possess the same electrophoretic mobility in various gel systems as the authentic proteins obtained from spinach thylakoid membranes. This discrepancy might be explainable if translation were to begin at another methionine codon within the open reading frames. If the fifth methionine residue (amino acid sixty, Figure 11) in the 51 kd, and the second methionine in the 44 kd apoprotein, which are both preceded by a Shine-Dalgarno type ribosomal binding sequence (GGAG; Shine and Dalgarno, 1974), were the first translational codon, the 51 kd protein would have an M of 49,865 and the 44 kd protein approximately 44 kd. However, the DNA sequence of at least the core of the P680 chlorophyll a apoprotein gene from Euoenothera plastome I shows 95% conservation of the sixty amino acids between the two

methionine residues although the nucleotide sequences differ by 8% (Morris et al., 1983). Nine of fourteen nucleotide differences between these two sequences are silent. This high degree of amino acid homology indicates that the full 508 amino acids are present in the primary translation product. A similar inconsistency between the size of a protein based on its electrophoretic mobility and that calculated from its deduced amino acid sequence has been noted for other hydrophobic proteins (Zurawski et al., 1982).

The 44 kd chlorophyll a apoprotein gene is preceded by a second open reading frame of 353 amino acid residues located on the same strand. This frame overlaps the 44 kd chlorophyll a apoprotein by 50 bp but is read in a different frame. The calculated molecular weight for this protein, inferred from the DNA sequences is 39,493 (Alt et al., 1984). The predicted amino acid sequence shows striking similarity to the "32 kd" protein involved in electron flow from PS II (Pfister et al., 1981; Zurawski et al., 1982; Satoh et al., 1983) and is very likely to code for the D2 polypeptide, a 34 kd component of PS II recently mapped on the Chlamydomonas reinhardtii chromosome (Rochaix, 1981) and sequenced (Rochaix et al., 1984). The order of homology between the two proteins is 25% (Alt et al., 1984) and can be seen by a comparison of the hydropathy profiles (Figure 37). These similarities suggest strongly that both polypeptides are folded in a similar manner in the membrane and play similar roles in photosynthetic electron transport.

The functional organization of the regions bearing these genes must be complicated. Recent investigations have revealed the presence of prokaryotic-like sequences for transcriptional regulation in the chloroplast DNA. In particular, promoter-like sequences (-10 and -35 regions) for the binding of RNA polymerase have been identified which show strong homologies with the E. coli promoter consensus sequence (for a review see Bohnert et al., 1982); stem-loop structures which function in E. coli as transcription termination signals or RNA processing sites have also been found in chloroplast DNA. Using the S1 nuclease technique, Shinozaki and Suguria (1982) have identified the 5' ends of the mRNA species for the beta-epsilon genes and the large subunit RUBP

carboxylase gene in tobacco. The ends are positioned just in front of (3') to the -10 and -35 bacterial-like promoter sequences providing proof that such signals function in chloroplasts.

Promoter sequences have been identified in the 5' flanking sequences of several plastid structural genes including the gene for the large subunit of ribulose 1,5-bisphosphate carboxylase in spinach (Zurawski et al., 1981) and tobacco (Shinozaki and Sugiura, 1982), the overlapping genes for the beta and epsilon subunits of the chloroplast ATP synthase (Krebbers et al., 1982; Shinozaki and Sugiura, 1982), and the gene for the 32 kd herbicide-binding protein in spinach and tobacco (Zurawski et al., 1982). In addition, E. coli extracts will transcribe (and translate) chloroplast genes yielding products identical with chloroplast isolated polypeptides (Bottomley and Whitfeld, 1979). Other prokaryotic-like transcriptional control mechanisms, which might function in the regulation of chloroplast gene expression include: (1) alteration of RNA polymerase promoter-recognition specificity by sigma-type factors, (2) positive or negative control through repressor/operator/inducer complexes, (3) transcription termination by attenuation, (4) read-through at termination signals via interaction of RNA polymerase with specific protein factors (anti-termination), (5) RNA processing and retro-regulation (Biro and Weissman, 1979; Rosenberg and Court, 1979) although none of these mechanisms have been amply demonstrated to exist in plastids.

A possible bacterial-like promotor sequence TAGAAT and its associated -35 region sequence TTGCGT, separated by 18 bp, are located 182 and 206 nucleotides, rerspectively, proximal to the first methionine of the P680 chlorophyll a apoprotein gene (Figure 11). Since the 44 kd chlorophyll a apoprotein gene and the D2 protein gene overlap, they are probably part of a single transcriptional unit. Sequence motifs reminiscent of prokaryotic promoter elements can be found in front of the D2 gene (Alt et al., 1984; Holschuh et al., 1984). The prokaryotic-like promoter for the 32 kd herbicide-binding polypeptide has been identified (Zurawski et al., 1982).

At the 3' end of the P680 chlorophyll a apoprotein gene, a thirty-

nine base pair stem-loop structure is located 295 bp distal to the TAA stop codon (Figure 12). Similarly, at the 3' end of the structural gene for the 44 kd chlorophyll a apoprotein a 22 bp sequence about 140 bp downstream of the TGA termination codon can form a stable hairpin ($\triangle G$=-29.6 kcal; Alt et al., 1984). Another stable stem-loop structure ($\triangle G$=-22.6 kcal) is present in the 3' flanking sequences of the D2 protein gene (which are the amino terminal sequences of the 44 kd apoprotein). A similar stable stem-loop structure has been found in the 3' flanking sequences of the 32 kd genes of tobacco and spinach (Zurawski et al., 1982). Similar stem-loop structures can be identified in the 3' untranslated regions of several other thylakoid membrane protein genes (review: Bohnert et al., 1982). In E. coli, such stem-loop structures have been identified as transcription termination sites (review: Rosenberg and Court, 1979) and in E. coli T7 phage as RNase III processing sites (Rosenberg and Kramer, 1977).

When chloroplast RNA is fractionated on denaturing agarose gels at least 100 distinct RNA species can be detected showing complementarity to 75% of the plastid chromosome (Poulsen, 1983; Heinemeyer et al., 1984; Alt et al., 1984). The genes for the P680 chlorophyll a apoprotein, cytochrome b_6 and subunit 4 map close together (Figure 17). Northern blot analysis shows several RNA species exist which are large enough to carry information for all three proteins (Figure 18). Smaller RNAs of 2.1 and 2.0 kb for the P680 chlorophyll a apoprotein, around 1.5 kb for the cytochrome b_6 plus subunit 4 and 0.9 kb for cytochrome b_6 alone, could arise through either alternate transcriptional activity or processing of larger transcripts. Nucleotide sequence analysis displays three possible transcriptional termination signals, one distal to the P680 chlorophyll a apoprotein gene (Figure 13), one between the genes for cytochrome b_6 and subunit 4 and one in the 3' flanking regions to the latter gene (Heinemeyer, 1983). Hence, the longest transcript beginning at the putative promotor of the PS II gene probably arises by read-through at two stem-loop structures. Separate promoters have not been identified for the genes of the cytochrome complex. However, stem-loop structures may also act as processing sites in a similar manner as

those identified in E. coli phages (Rosenberg and Kramer, 1977). All RNA species which are smaller than 5 kb but contain sequences complementary to all three genes could arise by RNA processing and intercistronic splicing (Morris et al., 1983; Heinemeyer et al., 1984). An exact size coincidence between RNA species from different segments of this DNA region is unlikely but has not been entirely ruled out.

Analyses performed since then with other clustered thylakoid membrane protein genes has shown that at least four regions of the plastid chromosome carrying the genes for the (1) P680 chlorophyll a apoprotein /cytochrome b_6/subunit 4, (Morris and Herrmann, 1984), (2) two P700 chlorophyll a apoproteins (Kirsch , personal communication), (3) D2 protein and 44 kd chlorophyll a apoprotein (Alt et al., 1984), (4) ATP synthase subunits proteolipid/CF_0-I/alpha (Westhoff et al., 1985) appear to be polycistronically transcribed. The half-life of the large RNAs is so long that they coexist with smaller RNA species for the individual genes. The steady-state concentrations of these RNA species, which could be achieved solely through posttranscriptional processing, may play a decisive role in controlling membrane biogenesis. The existence of split rRNA (Rochaix and Malnoe, 1978) and tRNA genes (e.g. Takaiwa and Sugiura, 1982) indicates that plastids do harbor the enzymatic equipment necessary for posttranscriptional endonucleolytic modification and splicing of such RNA classes.

The P680 chlorophyll a apoprotein and the 32 kd protein of PS II both contain stem-loop structures in the 5' flanking sequences to the structural genes (Figure 39; Zurawski et al., 1982). The S1 nuclease experiments on the P680 chlorophyll a apoprotein RNA disclosed a 5' end just to the right of structure No. 3 in Figure 39. Similar S1 nuclease analyses of the 32 kd gene in spinach show the 5' end of the 32 kd transcript is located just to the right of structure No. 4 in Figure 39 (Zurawski et al., 1982). The exact function of such structures in plastid DNA has not been clarified although it is known that in bacteria secondary structures in both DNA and RNA function as attenuators, terminators, protein-binding sites, and RNA processing sites.

(A)

-200

GCTATCAACTCCCATTGCGTATTGCTACTTATCGACTATAGAATAGATTTGTTTCTCTTTGTTCCTACAAACATA

-150 -100

ATTCTTCCATTATTACCAACAGAATAGAACAAACATTAACCCTTGTTCCGAGATAATCCATTGAACAAAGGATC

-50

CATTGCATAGTCATAGTCTTTTCCAATGCAATAAAGTTACATACTGTCATTTTTCTTTGATAAAGGGGTATTTCC

(B)

TTGACACGGGCATATAAGGCATGTTATACTGTTGAATAACAATC

Figure 39: Secondary structures in the 5' flanking regions to the (A) P680 chlorophyll a apoprotein gene, (B) 32 kd gene from spinach (Zurawski et al., 1982). The -35 and -10 sequences of the promoters are boxed and an arrow indicates the start site for transcription. Nucleotides which form the putative secondary structures are underlined.

The primary sequences of thylakoid membrane proteins deduced from the nucleotide sequence of their corresponding genes shows that the universal genetic code is used by higher plant plastomes (Whitfeld and Bottomley, 1983). A preference for A or T (70%) in the third position of degenerative codons has been observed in the P680 chlorophyll a apoprotein gene (Figure 14), the 44 kd chlorophyll a apoprotein gene (Table 4) and other plastid-coded thylakoid membrane proteins (Zurawski et al., 1982) which has been attributed to the low GC content of plastid DNA (38%; Herrmann et al., 1975; Herrmann and Possingham, 1980).

Analysis of the deduced primary sequences shows that approximately 50% of the amino acids of the two chlorophyll a binding apoproteins of PS II are hydrophobic as expected for membrane bound proteins. The hydropathy profiles for the 44 kd and the P680 chlorophyll a apoproteins of PS II show striking similarities to one another (Figures 15 and 28). Both have a large hydrophilic stretch in the carboxy terminal third of their molecules. By aligning the two chlorophyll a apoproteins for maximal homology (Figure 40), it can also be seen that the 44 kd chlorophyll a apoprotein has an amino terminal extension of approximately 30 amino acids when compared to the P680 chlorophyll a apoprotein and that a large segment of the latter protein located between transmembrane spans VI and VII accounts for most of the molecular weight difference between the two proteins. The overall homology between the two chlorophyll a apoproteins of PS II is on the order of 20% suggesting that they arose by gene duplication. Of the two chlorophyll a binding apoproteins of PS II, the 51 kd species is believed to carry both the P680 chlorophyll a and the pheophytin a molecule responsible for charge separation in PS II. The PS II reaction center is believed to contain from 5 to 40 chlorophyll a molecules per apoprotein chain or dimer (Bengis and Nelson, 1977; Chua et al., 1975; Delepelaire and Chua, 1979). If the chlorophyll a molecules are bound to histidine residues, then one might expect to see a difference between the distribution of such residues in the two chlorophyll a apoproteins. After aligning the P680 and 44 kd chlorophyll a apoprotein sequences (Figure 41) one sees that at least ten histidine residues are located in equivalent positions in the two chlorophyll a apoproteins. The P680 chlorophyll

a apoprotein has two unique histidine-cysteine-histidine sets of residues in spans III and IV which can be placed in juxtaposition within the membrane (Figure 16). The 44 kd has one set of such residues in span V (Figure 29). Both proteins also have a cysteine in the same position between segments IV and V.

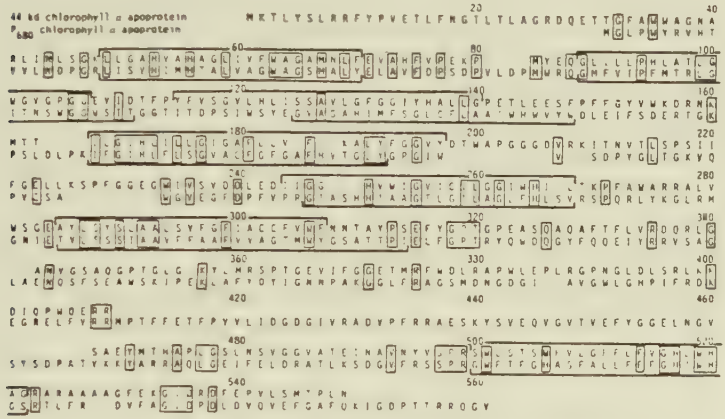


Figure 40: Comparison of the amino acid sequences derived from the nucleotide sequence of the 44 kd and 51 kd chlorophyll a apoproteins from spinach. Several amino acid gaps were allowed in order to maximize sequence homology. Homologous amino acid sequences are boxed and puta tive membrane spanning segments are indicated by horizontal brackets.

A more detailed picture of the homologous regions of these two proteins can be gained by tabulating the percentage of each of the amino acids found at the same position in the two proteins in comparison with the total number of such residues in the individual proteins (Table 5). Among the amino acids with uncharged polar side chains, those with hydroxyl groups show the least percent positional conservation among amino acids with 4% (threonine), 9-10% (serine) and 6-7% (tyrosine). Those having an amide group together with the nonpolar amino acids show an order of conservation ranging from valine (8-9%) to proline (26-28%) with the exception of tryptophan showing an exceptionally high 35%

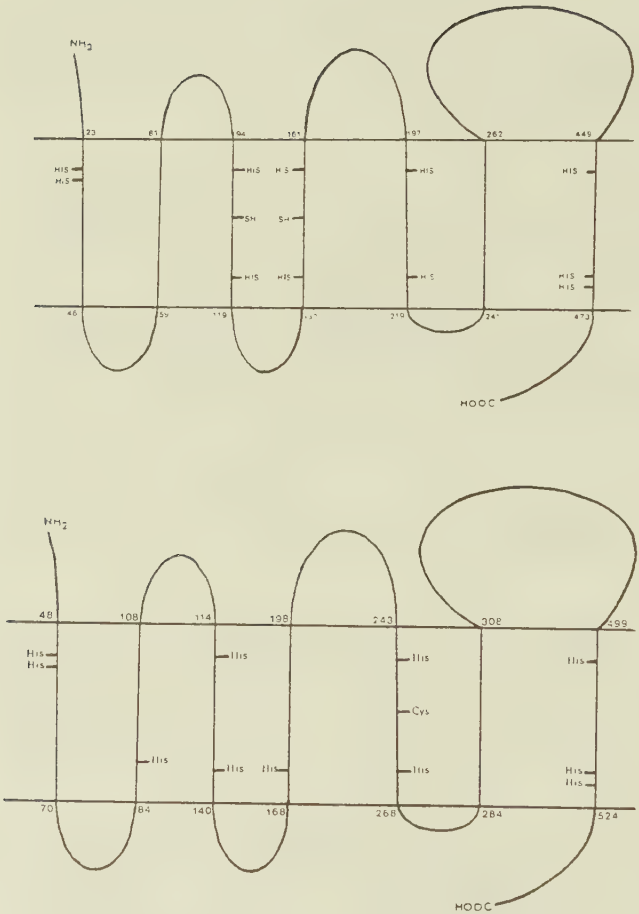


Figure 41: Comparison of positioning of histidine residues in the two chlorophyll a apoproteins of PS II. On the top is the P680 chlorophyll a apoprotein and on the bottom the 44 kd chlorophyll a apoprotein. Numbers indicate residues where the polypeptide enters or leaves the lipid bilayer.

positional conservation. More than half of the conserved glycine residues (33%) are found in membrane spanning regions where they probably disrupt the alpha helix configuration. In addition, all of the proline residues (except one), which are known to form kinks in alpha helices are found outside the membrane-spanning regions. Although the membrane-spanning regions of the proteins show a higher homology (40-60%) than the interspan regions, the individual hydrophobic amino acid residues show the least conservation of exact position. The most highly conserved amino acids are histidine (57-62%) and arginine (35-50%). Conserved arginine residues are found next to membrane-spanning sections where they probably serve to stabilize the membrane bound region; in the P680 chlorophyll a apoprotein fifteen residues are located in the hydrophilic section between membrane spans VI and VII of which eight residues are present in pairs. Five of these residues including one pair are conserved in the 44 kd apoprotein. This section of the apoproteins is believed to bulge into the lumen where it might associate with the watersplitting proteins.

The 32 kd herbicide-binding protein and the D2 protein of PS II also show significant homology to each other (26%) particularly in the three regions of the transmembrane segments III, IV, and V (50%). Both proteins also show significant homology with the L and M subunits of the R. capsulata reaction center (Youvan et al., 1984). Residues 180 to 219 of the D2 protein and 196 to 221 of the 32 kd protein also contain the sequence noted by Hearst and Sauer to be common to the L and M subunits of R. capsulata (1984); this sequence is characterized by the conserved sequence: pro-phe-his-met (or gly) and ala-met (or ile)-his-gly spaced so that the histidines are exactly 16 residues apart. The D2 protein like the M subunit, has one of these two histidine pairs whereas the 32 kd protein like the L subunit has both.

There are at least two regions of the 32 kd protein involved in herbicide binding. One is a tryptic peptide from residue 141 to 225 which binds azidoatrazine covalently (Gardner, 1981); the second is a serine residue at position 264 which in atrazine resistant mutants of Amaranthus hybridus (Hirschberg and McIntosh, 1984) and Solanum

TABLE 5: TABULATION OF % CONSERVED AMINO ACIDS BETWEEN THE P680 AND 44 KD CHLOROPHYLL A APOPROTEINS.

AMINO ACID	SAME POSITION[1]	44 KD[2] TOTAL	%[3]	51 KD TOTAL	%
ALANINE	6	40	15	39	15
ASPARTIC ACID	3	14	21	26	12
ASPARAGINE	2	14	14	8	25
ARGININE	6	12	50	17	35
CYSTEINE	1	4	25	3	33
GLUTAMIC ACID	3	22	14	22	14
GLUTAMINE	2	8	25	13	15
GLYCINE	20	59	33	60	33
HISTIDINE	8	13	62	14	57
ISOLEUCINE	5	26	19	30	17
LEUCINE	11	57	19	39	28
LYSINE	3	13	23	13	23
METHIONINE	2	8	25	11	18
PHENYLALANINE	5	33	15	41	12
PROLINE	7	25	28	27	26
SERINE	3	27	11	34	9
THREONINE	1	25	4	25	4
TRYPTOPHAN	6	17	35	17	35
TYROSINE	1	15	7	17	6
VALINE	3	32	9	40	8

1 Number of particular amino acid located at the same position (Figure 39).

2 Total number of particular amino acid in the designated protein.

3 Number of amino acid at same position divided by total number of particular amino acid in the protein X 100.

nigrum (Hirschberg et al., 1984) is changed to glycine. The D2 protein has a serine at position 263 and is probably the 34 kd atrazine binding component found by Steinback et al., 1981.

The sequence of the 32 kd gene from the herbicide-resistant strain of Amaranthus retroflexus shows a single amino acid change from serine to glycine at position 264. The sequence of the wildtype 32 kd gene from spinach, tobacco, (Zurawski et al., 1982), soybean (Spielmann and Stutz, 1983), mustard (Link and Langridge, 1984), and Amaranthus hybridus (Hirschberg and McIntosh, 1984) all contain a serine residue at this position in the amino acid sequence suggesting that the change to glycine could be reponsible for herbicide resistance. Triazine herbicides are believed to bind to the Q_B site on the 32 kd protein thereby interfering with electron transport (Pfister et al., 1981). The P680 chlorophyll a apoprotein is believed to contain the binding site for Q_A. The membrane-internal, nine amino acid sequence: his-leu-phe-leu-ser-gly-val-ala-cys-, starting at residue 142 of the P680 chlorophyll a apoprotein, is nearly identical to an eight residue sequence: his-phe-leu-leu-gly-val-ala-cys- from amino acid 118 in the 32 kd herbicide binding protein of spinach (Zurawski et al., 1982). Such conserved sequences are also potential candidates for the plastoquinone binding sites, Q_A and Q_B, respectively.

If the endosymbiotic theory for the origin of chloroplasts is correct one expects to see similarites between photosynthetic bacteria and chloroplast electron transport systems. The recent finding that two of the structural polypeptides (L and M of the Rhodopseudomonas reaction center) are homologous with two of the proteins belonging to PS II of the chloroplast electron transport chain provides strong support for the evolutionary lineage : purple nonsulfur bacteria to cyanobacteria to chloroplasts (Youvan et al., 1984; Youvan and Marrs, 1984).

In the Z scheme of electron transport (Figure 1), cytochrome b_6 is depicted as an intermediate between the pool of plastoquinones and the Rieske FeS protein. The recent discovery that cytochrome b_6 is associated with subunit 4, the Rieske FeS protein and cytochrome

f in a membrane complex (Hurt and Hauska, 1981) suggests that these electron carriers are indeed the exact counterparts of the purple nonsulfur bacterial photosynthetic complex. The fbc operon encoding the FeS protein, cytochrome b and cytochrome c_1 has recently been identified in R. sphaeroides (Gabellini et al., 1985). The electron carriers between PS I and PS II also show striking similarity to the respiratory electron transport complex of mitochondria. The amino acid sequences of cytochrome b_6 of chloroplasts and cytochrome b of mitochondria have recently been compared (Heinemeyer, 1983; Widger et al., 1984). Two conserved pairs of histidine residues separated by 13 and 14 amino acids were identified as the probable heme ligands. The amino acid sequence of cytochrome f has been deduced from the DNA sequence of the gene in pea (Willey et al., 1984) and spinach (Alt and Herrmann, 1984). The heme-binding site, cys-ala-asn-cys-his is similar to that of cytochrome c_1: cys-ser-ser-cys-his (Wakabayashi et al., 1982).

The functional analogy of the cytochrome b/f complex of chloroplasts and the cytochrome bc_1 complex of mitochondria would suggest that cytochrome b_6 functions in normal linear electron transport between the plastoquinone pool and cytochrome f. However, in analogy with the cyclic system in purple nonsulfur bacteria, cytochrome b_6 would receive electrons from the quinone pool and pass them to an FeS protein. The FeS protein, e.g. Rieske factor, would communicate with a c-type cytochrome, e.g. cytochrome f.

The homologies between the reaction centers of purple nonsulfur bacteria and chloroplast PS II give important clues to the functional roles of the various proteins in binding carotenoid, chlorophyll and quinone cofactors. The reaction center of R.sphaeroides is currently being subjected to X-ray crystallographic analysis. These data on the purple nonsulfur bacterium should provide a mechanistic explanation for photosynthetic electron transport that is also applicable to higher plants.

7.0 Summary

The structures of two genes coding for reaction center proteins of photosystem II from spinach (Spinicia oleracea) have been investigated using DNA cloning and sequencing techniques: the P680 chlorophyll a apoprotein, which catalyses the primary light reaction, and the 44 kd chlorophyll a binding apoprotein.

Prokaryotic-like transcription promoter and terminator signals are located in the 5' and 3' flanking regions, respectively, of the P680 apoprotein structural gene. Seven discrete RNA species ranging in size from 2.0 to over 5.0 kilobases display complementarity to apoprotein coding sequences implying that the region can be polycistronically transcribed. The larger transcripts include information for cytochrome b_6 and subunit 4 proteins of the cytochrome b/f complex.

The amino acid sequences of the P680 chlorophyll a apoprotein and the 44 kd chlorophyll a apoprotein were deduced from the DNA sequences. The largest open reading frames code for proteins of M_r 56,246 and 51,785 containing 508 and 473 amino acid residues, respectively. A comparison of the two chlorophyll a apoproteins along their entire lengths shows about 20% homology suggesting that they may have arisen by gene duplication. Both apoproteins possess a high content of hydrophobic amino acids clustered into seven core regions which probably traverse the membrane; both also contain a hydrophilic domain of between 150 and 200 amino acids interspersed between transmembrane spans VI and VII. Spanning segments III, IV, V and VII of the P680 chlorophyll a apoprotein and III, V, and VII of the 44 kd chlorophyll a apoprotein each contain a pair of histidine residues separated by 13 or 14 amino acids. These residues may be involved in the binding of chlorophyll a. Specific functional roles can also be expected for the highly conserved arginine (35 to 50%), tryptophan (35%) and cysteine (25 to 33%) residues.

Triazine herbicides are believed to disrupt electron transport by binding to a quinone site on the 32 kd protein of the reducing side of PS II. The gene for the 32 kd protein from a triazine-resistant strain of Amaranthus retroflexus was cloned and sequenced. The structural gene contains a single A to G transversion which causes a replacement of

serine by glycine at position 264 in the amino acid sequence. Since serine is present at this position in the wild type 32 kd protein from five different species of plants, and a similar glycine replacement has been found in another triazine-resistant species of Amaranthus, the replacement is very likely to be responsible for herbicide resistance.

8.0 Literature

Alt,J., Herrmann, R.G.
Nucleotide sequence of the gene for the pre-apocytochrome f in the spinach plastid chromosome.
Current Genet. 8, 551-557 (1984).

Alt, J., Morris, J., Westhoff, P., Herrmann, R.G.
Nucleotide sequence of the clustered genes for the 44 kd chlorophyll a apoprotein and the "32 kd"-like protein of the photosystem II reaction center in the spinach plastid chromosome.
Current Genet. 8, 597-606 (1984).

Alt,J., Winter, P., Sebald, W., Moser, J.G., Schedel, R., Westhoff, P., Herrmann, R.G.
Localization and nucleotide sequence of the gene for the ATP-synthase proteolipid subunit on the spinach plastid chromosome.
Current Genet. 7, 129-138 (1983).

Alt, J., Westhoff, P., Sears, B.B., Nelson, N., Hurt, E., Hauska, G., Herrmann, R.G.
Genes and transcripts for the polypeptides of the cytochrome b_6/f complex from spinach thylakoid membranes.
EMBO J. 2, 979-986 (1982).

Akerlund, H.E., Jannsson, C., Andersson, B.
Reconstitution of photosynthetic water splitting in inside-out thylakoid vesicles and identification of a participating polypeptide.
Biochim. Biophys. Acta 681, 1-10 (1982).

Anderson, J.M., Andersson, B.
The architecture of photosynthetic membranes: lateral and transverse organization.
TIBS, August, 288-292 (1982).

Anderson, J.M., Barrett, J., Thorne, S.W.
In: Photosynthesis (Akoyunoglou, G., ed.) Vol. III, pp. 301-315, (1981)
Balaban International Science Services, Philadelphia.

Appleyard, R.K.
Segregation of new lysogenic types during growth of a doubly lysogenic strain derived from Escherichia coli K12.
Genetics 39, 440-452 (1954).

Bachmann, R.C., Gillies, K., Takemoto, T.Y.
Membrane topology of the photosynthetic reaction center polypeptides of Rhodopseudomonas sphaeroides.
Biochemistry 20, 4590-4596 (1981).

Bengis, C., Nelson, N.
Purification and properties of the PS I reaction center from chloroplasts.
J. Biol. Chem. 250, 2783-2788 (1975).

Bengis, C., Nelson, N.
Subunit structure of chloroplast PS I reaction center.
J. Biol. Chem. 252, 4564-4569 (1977).

Biro, P.A., Weissman, S.M.
Recognition and control sequences in nucleic acids.
In: Molecular genetics, Part III (Stent, Ed.) pp. 177-245 (1979),
Academic Press Inc., New York.

Birnboim, H.C. Doly, J.
A rapid alkaline extraction procedure for screening recombinant plasmid DNA.
Nucleic Acids Res. 7, 1513-1520 (1979).

Bohnert, H.J., Crouse, E.J., Schmitt, J.M.
Organization and expression of plastid genomes.
In: Encycl. of Plant Physiology, New series, Vol. 14B (Parthier, B., Boulter, D., eds.) pp. 477-530 (1982). Springer-Verlag, Berlin Heidelberg.

Bolivar, F., Rodriguez, R., Greene, P.J., Betlach, M., Heyneker, H.L., Boyer, H.W., Crosa, J., Falkow, S.
Construction and characterisation of new cloning vehicles II; a multiple cloning system II.
Gene 2, 95-113 (1977).

Bolivar, F.
Construction and characterisation of new cloning vehicles III: derivatives of plasmid pBR322 carrying unique Eco RI sites for selection of Eco RI generated recombinant molecules.
Gene 4, 121-136 (1978).

Bonner, T.I., Brenner, B.R., Neufeld, B.R., Britten, R.J.
Reduction in the rate of DNA reassociation by sequence divergence.
J. Mol. Biol. 81, 123 (1973).

Bottomley, W., Whitfeld, P.
Cell-free transcription and translation of total spinach chloroplast DNA.
Eur. J. Biochem. 93, 31-39 (1979).

Bünemann, H., Westhoff, P., Herrmann, R.G.
Immobilization of denatured DNA to macroporous supports I. Efficiency of different coupling procedures.
Nucleic Acids Res. 10, 7163-7180 (1982).

Camm, E.L. Green, B.R.
Relationship between the two minow chlorophyll a-protein complexes and the photosystem II reaction center.
Biochim. Biophys. Acta 724, 291-293 (1983).

Casey, J., Davidson, N.
Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide.
Nucleic Acids Res. 4, 1539 (1977).

Chua, N.H., Matlin, K., Bennoun, P.
A chlorophyll-protein complex lacking in photosystem I mutants of Chlamydomonas reinhardii.
J. Cell Biol. 67, 361-377 (1975).

Clothia, C.
The nature of accessible and buried surfaces in proteins.
J. Mol. Biol. 105, 1-14 (1976).

Clewell, D.B.
Nature of Col E_1 plasmid replication in Escherichia coli in the presence of chloramphenicol.
J. Bacteriol. 110, 667 (1972).

Cox, R.P., Andersson, B.
Lateral and transverse organization of cytochromes in the chloroplast thylakoid membrane.
Biochem. Biophys. Res. Commun., 103, 1336-1342 (1981).

Curtis, S.E., Haselkorn, R.
Isolation and sequence of the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase from the cyanobacterium Anabaena 7120.
Proc. Natl. Acad. Sci. USA 80, 1835-1839 (1983).

Dagert, M., Ehrlich, S.D.
Prolonged incubation in calcium chloride improves the competence of Escherichia coli cells.
Gene 6, 23-28 (1979).

Delepelaire, P., Chua, N.H.
Lithium dodecyl sulfate/-polyacrilamide gel electrophoresis of thylakoid membranes at 4°C: Characterizations of two additional chlorophyll a-protein complexes.
Proc. Natl. Acad. Sci. USA 76, 111-115 (1979).

Deno, H., Shinozaki, K., Sugiura, M.
Nucleotide sequence of tobacco chloroplast gene for the alpha subunit of proton-translocating ATPase.
Nucleic Acids Res. 11, 2185-2191 (1983).

Dove, W.F., Davidson, N.
Cation effects on the denaturation of DNA.
J. Mol. Biol. 5, 467-486 (1972).

Driesel, A.J., Speirs, J., Bohnert, H.J.
Spinach chloroplast mRNA for a 32000 dalton polypeptide: Size and localization on the physical map of the chloroplast DNA.
Biochim. Biophys. Acta 610, 297-310 (1980).

Dron, M., Rahine, M., Rochaix, J.D.
Sequence of the chloroplast DNA region of Chlamydomonas reinhardii containing the gene of the large subunit of ribulose bisphosphate carobxylase and parts of its flanking genes.
M. Mol. Biol. 162, 775-793 (1982).

Edwards, K. Kössel, H.
The rRNA operon from Zea mays chloroplasts: nucleotide sequence of 23S rDNA and its homology with E. coli 23S rDNA.
Nucleic Acids Res. 9, 2853-2869 (1981).

Gabellini, N. Harnisch, U., McCarthy, J.F.G., Hauska, G. Sebald, W.
Cloning and expression of the fbc operon encoding the FeS protein, cytochrome b and cytochrome c_1 from the Rhodopseudomonas sphaeroides b/c_1 complex.
EMBO J. 4, 549-553 (1985).

Gauss, D.H., Spinzl, M.
Compilation of sequences of tRNA genes.
Nucleic Acids Res. 11, 1-53 (1983a)

Gardner, G.
Azidoatrazine: Photoaffinity label for the site of triazine herbicide action in chloroplasts.
Science 211, 937-940 (1981).

Gay, N.J., Walker, J.E.
The ATP operon: nucleotide sequence of the promoter and the genes for the membrane proteins, and the delta subunit of Escherichia coli ATP-synthase.
Nucleic Acids Res 9, 3919-3926 (1981).

Grunstein, M., Hogness, D.
Colony hybridization: A method for the isolation of cloned DNAs that contain a specific gene.
Proc. Natl. Acad. Sci. 72, 3961-3965 (1975).

Heinemeyer, W.
Nukleotidsequenzierungen an Genen des Cytochrom b_6/f-Komplexes aus Spinat.
Diplomarbeit, Universität Düsseldorf, 1983.

Heinemeyer, W., Alt, J., Herrmann, R.G.
Nucleotide sequence of the clustered genes for apocytochrome b_6 and subunit 4 of the cytochrome b/f complex in the spinach plastid chromosome.
Current Genetics 8, 543-549 (1984).

Hensgens, L.A.M., Grivell, L.A., Borst, P., Bos, J.L.
Nucleotide sequence of the mitochondrial structural gene for subunit 9 of yeast ATPase complex.
Proc. Natl. Acad. Sci. 76, 1663-1667 (1979).

Herrmann, R.G.
The preparation of circular DNA from plastids.
In: Edelman, M., Hallick, R., Chua, N.H. (eds.) Methods in chloroplast molecular biology, pp. 259-280 (1982) Amsterdam-Elsevier/North-Holland Press.

Herrmann, R.G., Bohnert, H.J., Kowallik, K.V., Schmitt, J.M.
Size, conformation and purity of chloroplast DNA of some higher plants.
Biochim. Biophys. Acta 378, 305-317 (1975).

Herrmann, R.G., Possingham, J.V.
Plastid DNA-the plastome.
In: Reinert, J. (ed.) Results and problems in cell differentiation, Vol. 10, pp 454-96. (1980) Berlin-Heidelberg-New York:Springer Verlag.

Herrmann, R.G., Whitfeld, P.R.
Restriction mapping of chloroplast DNA using low-melting temperature agarose.
In: Edelman, M., Hallick, R., Chua, N.H. (eds.) Methods in chloroplast molecular biology. pp 451-468. (1982). Amsterdam:Elsevier/North-Holland Press.

Herrmann, R.G., Seyer, P. Schedel, R., Gordon, K., Bisanz, C., Winter, P., Hildebrandt, J.W., Wlaschek, M., Alt, J. Driesel, A.J., Sears, B.B.
The plastid chromosomes of several dicotyledons.
In: Bücher, T., Sebald, W., Weiss, H. (eds.) Biological chemistry of organelle formation, pp 97-112 (1980). Berlin-Heidelberg-New York: Springer Verlag.

Herrmann, R.G., Westhoff, P., Alt, J., Winter, P., Tittgen, JH., Bisanz, C., Sears, B.B., Nelson, N., Hurt, E., Hauska, G., Viebrock, A., Sebald, W.
Identification and characterization of genes for polypeptides of the thylakoid membrane.
In: Cifferi, O. (ed.) Structure and function of plant genomes; pp 143-154 (1983).Academic Press, New York.

Herrmann, R.G., Westhoff, P., Alt, J., Tittgen, J., Nelson, N.
Thylakoid membrane proteins and their genes.
Manuscript in preparation (1985).

Hearst, J.D. Sauer, K.
Protein sequence homologies between portions of the L and M subunits of reaction centers of Rhodopseudomonas capsulata and the Q_B-protein of chloroplast thylakoid membranes: a proposed relation to quinone-binding sites.
Z. Naturforsch. 39c, 421-424 (1984).

Hill, R., Bendall, F.
Function of the two cytochrome components in chloroplasts: a working hypothesis.
Nature Lond., 186, 136-137 (1960).

Hirschberg, J., Bleecker, A., Kyle, D.J., McIntosh, L., Arntzen, C.J.
The molecular basis of triazine-herbicide resistance in higher-plant chloroplasts.
Z. Naturforsch. 39c, 412-420 (1984).

Hirschberg, J., McIntosh, L.
Molecular basis of herbicide resistance in Amaranthus hybridus.
Science 222, 1346-1348 (1983).

Holschuh, K., Bottomley, W., Whitfeld, P.R.
Structure of the spinach chloroplast genes for the D2 and 44 kd reaction-centre proteins of photosystem II and for $tRNA^{Ser}$ (UGA).
Nucleic Acids Res. 1 , 8819-8833 (1984).

Howe, C.J., Auffret, A.D., Doherty, A., Bowman, C.M., Dyer, T.A., Gray, J.C.
Location and nucleotide sequence of the gene for the proton-translocating subunit of wheat chloroplast ATP synthase.
Proc. Natl. Acad. Sci. USA 79, 6903-6907 (1982).

Hurt,E. Hauska, G.
A cytochrome b_6/f complex of five polypeptides with plastoquinol-plastocyanin-oxidoreductase activity from spinach chloroplasts.
Eur. J. Biochem. 117, 591-599 (1981).

Hurt, E., Hauska, G.
Identification of the polypeptides in the cytochrome b_6/f-complex from spinach chloroplasts with redox center carrying subunits.
J. Bioenerg. Biomem. 14, 405-424 (1982).

Hutton, J.A.
Renaturation kinetics and thermostability of DNA in aqueous solutions of formamide and urea.
Nucleic Acids Res. 4, 3537-3555 (1977).

Jacobsen, H., Klanow, H., Ovargaard-Hansen, K.
The N-terminal amino-acid sequences of DNA polymerase I from Escherichia coli and of the large and the small fragments obtained by a limited proteolysis.
Eur. J. Biochem. 45, 623 (1974).

Jay, F., Lambillotte, M., Stark, W., Mühlethaler, K.
The preparation and characterisation of native photoreceptor units from the thylakoids of Rhodopseudomonas viridis.
EMBO J. 3, 773-776 (1984).

Kaplan, S., Arntzen, C.J.
Photosynthetic membrane structure and function.
In: Photosynthesis, Vol 1, Energy conversion by plants and bacteria (Govindjee, ed.), pp 65-151 (1982). Academic Press, New York.

Katz, L. Kingsbury, D.T., Helinski, D.R.
Stimulation by cyclic adenosine monophosphate of plasmid deoxyribonucleic acid replication and catabolite repression of the plasmid deoxyribonucleic acid-protein relaxation complex.
J. Bacteriol. 114, 577-591 (1973).

Klimov, V.V., Klevanik, A.A., Shuvalow, V.A., Krasnovsky, A.A.
Reduction of pheophytin in the primary light reaction of photosystem II.
FEBS Lett. 82, 183-186 (1977).

Kröger, M., Kröger-Block, A.
A flexible new computer program for handling DNA sequence data.
Nucleic Acids Res. 10, 229-236 (1982).

Krebbers, E.T., Larrinua, I.M., McIntosh, L., Bogorad, L.
The maize chloroplast genes for the beta and epsilon subunits of photosynthetic coupling factor CF_1 are fused.
Nucleic Acids Res. 10, 4985-5002 (1982).

Kyte, J., Doolittle, F.
A simple method for displaying the hydropathic character of a protein.
J. Mol. Biol. 157, 105-132 (1982).

Lehrach, H., Diamond, D., Wozney, J.M., Boedtker, H.
RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination.
Biochemistry 16, 4743-4751 (1977).

Link, G., Langridge, U.
Structure of the chloroplast gene for the precursor of the M_r 32,000 photosystem II protein from mustard (Sinapis alba L.).
Nucleic Acids Res. 12, 945-960 (1984).

Lis, J.T.
Fractionation of DNA fragments by polyethylene glycol induced precipitation.
In: Methods in Enzymology 65, 347-353 (1980).

Maniatis, T., Jeffrey, A., Kleid, D.G.
Nucleotide sequence of the rightward operator of phage lambda.
Proc. Natl. Acad. Sci. 72, 1184 (1975).

Maniatis, T., Fritsch, E.F., Sambrook, J.
Molecular cloning, a laboratory manual.
Cold Spring Harbor Laboratory, 1982.

Maxam, A.M., Gilbert, W.
A new method for sequencing DNA
Proc. Natl. Acad. Sci. USA 74, 560-564 (1977).

McConaughy, B.L., Laird, C.D., McCarthy, B.J.
Nucleic acid reassociation in formamide.
Biochemistry 8, 3289 (1969).

McIntosh, L., Poulsen, C., Bogorad, L.
Chloroplast gene sequence for the large subunit of ribulose bisphosphate carboxylase of maize.
Nature 288, 556-569 (1980).

Metz, J.G., Miles, D.
Use of a nuclear mutant of maize to identify components of photosystem II.
Biochim. Biophys. Acta 681, 96-102 (1982).

Morris, J., Herrmann, G.
Nucleotide sequence of the gene for the P680 chlorophyll a apoprotein of the photosystem II reaction center from spinach.
Nucleic Acids Res. 12, 2837-2850 (1984).

Morris, J., Offermann, K., Herrmann, R.G.
Nucleotide sequence of the gene for the 51 kd chlorophyll a binding protein associated with photosystem II reaction center in spinach, Oenothera hookeri and argillicola.
Abstracts, Sixth Int. Congress on Photosynthesis, Vol 2, p. 321 (1983).

Mullet, J.E., Arntzen, C.J.
Identification of a 32-34 kd polypeptide as a herbicide acceptor protein in photosystem II.
Biochim. Biophys. Acta 635, 236-248 (1981).

Mullet, J.E., Burke, J.E., Arntzen, C.J.
Chlorophyll proteins of photosystem I.
Plant Physiol. 65, 814-822 (1980).

Nabedryk, E., Tiede, D.M., Durron, P.L., Breton, J.
Conformation and orientation of the protein in the bacterial photosynthetic reaction center.
Biochim. Biophys. Acta 682, 273-280 (1982).

Nakatani, H.Y.
In: The oxygen evolving system of photosynthesis. (Inoue, Y., Crofts, A.R., Govindjee, M.N., Renger, G., Satoh, K., eds.) pp 201-208 (1983). Academic Press, Tokyo.

Nechushtai, R., Nelson, N., Mattoo, A.K., Edelman, M.
Site of synthesis of subunits to photosystem I reaction center and the proton-ATPase in Spirodela.
FEBS Lett. 125, 115-119 (1981).

Nelson, N.
Proton-ATPase of chloroplasts.
In: Current topics in bioenergetics. Sanadi, D.R. (ed.)
Vol 11, pp. 1-33, (1981) Academic Press, New York.

Nelson, N., Notsani, B.B.
Function and organization of individual polypeptides in chloroplast photosystem I reaction center.
In: Bioenergetics of Membranes. Packer, L., Papageorgiou, G.C., Trebst, A. (eds.) pp 233-244. Amsterdam: Elsevier/North-Holland Press.

Okamura, M.Y., Steiner, L.A., Feher, G.
Characterization of reaction centers from photosynthetic bacteria-I. Subunit structure of the protein mediating the primary photochemistry in Thodopseudomonas sphaeroides R-26.
Biochem. 13, 1394-1402 (1974).

Parthier, B.
The cooperation of nuclear and plastid genomes in plastid biogenesis and differentiation.
Biochem. Physiol. Pflanzen 177, 283-317 (1982).

Pfister, K., Steinback, K.E., Gardner, G., Arntzen, C.J.
Identification of the photosystem II herbicide binding protein.
Plant Physiol. Suppl. 65, 10 (1980).

Phillips, A.L., Gray, J.C.
Location and nucleotide sequence of the gene for the 15.2 kD polypeptide of the catochrome b-f complex from pea chloroplasts.
Mol. Gen Genet 194, 477-484 (1984).

Poulsen, C.
The barley chloroplast genome: physical structure and transcriptional activity in vivo.
Carlsberg Res. Commun. 48, 57-80 (1983).

Pribnow, D.
Bacteriophage T7 early promoters: nucleotide sequences of two RNA polymerase binding sites.
J. Mol. Biol. 99, 419-443 (1975).

Rigby, P.W.J., Dieckmann, M., Rhodes, C., Berg, P.
Labelling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I.
J. Mol. Biol. 113, 237-251 (1977).

Rochaix, J.D.
Organization, function and expression of chloroplast DNA of Chlamydomonas reinhardii.
Experientia 37, 323-340 (1981).

Rochaix, J.D., Dron, M., Rahire, M., Malnoe, P.
Sequence homology between the 32 K dalton and the D2 chloroplast membrane polypeptides of Chlamydomonas reinhardii.
Plant Mol. Biol. 3, 363-370 (1984).

Rochaix, J.D., Malnoe, P.
Anatomy of the chloroplast ribosomal DNA of Chlamydomonas reinhardii
Cell 15, 661-670 (1981).

Rosenberg, M., Court, D.
Regulatory sequences involved in the promotion and termination of RNA transcription.
Annual Rev. Genetics 13, 319-353 (1979),

Rosenberg, M., Kramer, R.A.
Nucleotide sequence surrounding a ribonuclease III processing site in bacteriophage T7 RNA.
Proc. Natl. Acad. Sci. USA 74, 984-988 (1977).

Schmidt, G.W., Bartlett, S.G., Grossman, A.R., Cashmore, A.R., Chua, N.H.
Biosynthetic pathways of two polypeptide subunits of the light-harvesting chlorophyll a/b protein complex.
J. Cell Biol. 91, 468-478 (1981)

Schwarz, Z., Jolly, S.O., Steinmety, A., Bogorad, L.
Overlapping divergent genes in the maize chloroplast chromosome and in vitro transcription of the gene for $tRNA^{HIS}$.
Proc. Natl. Acad. Sci. USA 78, 4748-4752 (1981).

Schwarz, Z. Kössel, H.
The primary structure of 16S rDNA from Zea mays chloroplast is homologous to E. coli 16S rRNA.
Nature (London) 283, 739-742 (1980).

Shine, J., Dalgarno, L.
The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: Complementarity of nonsense triplets and ribosome binding sites.
Proc. Natl. Acad. Sci. USA 71, 1342-1346 (1974).

Shinozaki, K., Deno, H., Kato, A., Sugiura, M.
Overlap and cotranscription of the genes for the beta and epsilon subunits of tobacco chloroplast ATPase.
Gene 24, 147-155 (1983).

Shinozaki, K., Sugiura, M.
Sequence of the intercistronic region between the ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit and the coupling factor beta subunit gene.
Nucleic Acids Res. 10, 4923-4934.

Shinozaki, K., Sugiura, M.
The nucleotide sequence of the tobacco chloroplast gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase.
Gene 20, 91-102 (1982)

Shinozaki, K. Yamada, C., Takahata, N. Sugiura, M.
Molecular cloning and sequence analysis of the cyanobacterial gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase.
Proc. Natl. Acad. Sci. USA 80, 4050-4054 (1983).

Southern, E.M.
Detection of specific sequences among DNA fragments separated by gel electrophoresis.
J. Mol. Biol. 98, 503-517 (1975).

Spielmann, A., Stutz, E.
Nucleotide sequence of soybean chloroplast DNA regions which contain the psbA and trnH genes and cover the ends of the large single copy region and one end of the inverted repeats.
Nucleic Acids Res. 11, 7157-7167 (1983).

Steinback, K.E., McIntosh, L., Bogorad, L., Arntzen, C.J.
Identification of the triazine receptor protein as a chloroplast gene product.
Proc. Natl. Acad. Sci. USA 78, 7463-7467 (1981).

Steiner, L.A., Lkamura, M.Y., Lopes, A.D., Moskovitz, E., Feher, G.
Characterization of reaction centers from photosynthetic bacteria.
II. Amino acid composition of the reaction center protein and its subunits in Rhodopseudomonas spheroides R-26.
Biochemistry 13, 1403-1410 (1974).

Sutton, M.R., Rosen, D., Feher, G., Steiner, L.A.
Amino-terminal sequences of the L, M, and H subunits of reaction centers from the photosynthetic bacterium Rhodopseudomonas sphaeroides R-26.
Biochemistry 21, 3842-3849 (1982).

Takaiwa, F., Sugiura, M.
Nucleotide sequence of the 16S-23S spacer region in an rRNA gene cluster from tobacco chloroplast DNA.
Nucleic Acids Res. 10, 2665-2675 (1982).

Thomas, P.S.
Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose.
Proc. Natl. Acad. Sci. USA 77, 5201-5205 (1980).

Tinoco, I. Jr., Borer, P.N., Dengler, B., Levine, M.D.
Improved estimation of secondary structure in ribonucleic acids.
Nature New Biology 246, 40-41 (1973).

Valkirs, G.G. Feher, G.
Topography of reaction center subunits in the membrane of the photosynthetic bacterium, SRhodopseudomonas sphaeroides.
J. Cell Biol. 95, 179-188 (1982).

Wahl, G.M., Stern, M., Stark, G.R.
Efficient transfer of large DNA fragments from agarose gels to diazobenyloxymethal-paper and rapid hybridization by using dextran sulfate.
Proc. Natl. Acad. Sci. 76, 3683-3697 (1979).

Wakabayashi, S., Matsubara, H., Kim, C.H., King, T.R.
Structural studies of bovine heart cytochrome c_1.
J. Biol. Chem. 257, 9335-9344 (1982).

Westhoff, P., Alt, J., Herrmann, R.G.
Localization of the genes for the two chlorophyll a-conjugated polypeptides (mol wt. 51 and 44 kd) of the photosystem II reaction center on the spinach plastid chromosome.
EMBO J. 2, 2229-2237 (1983).

Westhoff, P., Alt, J., Nelson, N., Bottomley, W., Bünemann, H., Herrmann, R.G.
Genes and transcripts for the P700 chlorophyll a apoprotein and subunit 2 of the photosystem I reaction center complex from spinach thylakoid membranes.
Plant Molecular Biol. 2, 95-107 (1983).

Westhoff, P., Alt, J., Nelson, N., Herrmann, R.G.
Genes and transcripts for the ATP synthase CF_0 subunits I and II from spinach thylakoid membranes.
submitted to Mol. Gen. Genet. (1985).

Westhoff, P., Bartlett, S.G., Jansson, C., Klein-Hitpaß, L. Larsson, C. Berzborn, R., Herrmann, R.G.
Coding sites and uptake into chloroplasts of polypeptides associated with the water-splitting apparatus of photosystem II in Spinicia oleracea.
Abstracts, Sixth International Congress on Photosynthesis Vol. 2 p. 330 (1983).

Westhoff, P., Nelson, N., Bünemann, H., Herrmann, R.G.
Localization of genes for coupling factor subunits on the spinach plastid chromosome.
Current Genet. 4, 109-120 (1981).

Whitfeld, P.R., Bottomley, W.
Mapping of the gene for the large subunit of ribulose bisphosphate carboxylase on spinach chloroplast DNA.
Biochem. Int 1, 172-178 (1980).

Whitfeld, P.R., Bottomley, W.
Organization and structure of chloroplast genes.
Ann. Rev. Plant Physiol 34, 279-310 (1983).

Widger, W.R., Cramer, W.A., Herrmann, R.G., Trebst, A.
Sequence homology and structural similarily between the b cytochrome of mitochondrial complex III and the chloroplast b_6f complex: position of the cytochrome b hemes in the membrane.
Proc. Natl. Acad. Sci. 81, 674-678 (1984).

Willey, D.L., Auffret, A.D., Gray, J.C.
Structure and topology of cytochrome f in pea chloroplast membranes.
Cell Vol. 36, 555-562 (1984).

Willey, D.L., Howe, C.J., Auffret, A.D., Bowman, C.M., Dyer, T.A., Gray, J.C.
Location and nucleotide sequence of the gene for cytochrome f in wheat chloroplast DNA.
Mol. Gen. Genet. 194, 416-422 (1984).

Willey, D.L., Huttly, A.K., Phillips, A.L., Gray, J.C.
Localization of the gene for cytochrome f in pea chloroplast DNA.
Mol. Gen. Genet. 189, 85-89 (1983).

Williams, J.C., Steiner, L.A., Ogden, R.C., Simon, M.I., Feher, G.
Primary structure of the M subunit of the reaction center from Rhodopseudomonas sphaeroides.
Proc. Natl. Acad. Sci. USA 80, 6505-6509 (1983).

Wlaschek, M.
Diplomarbeit, Universität Düsseldorf (1982).

Wolfenden et al.,
Affinities of amino acid side chains for solvent water.
Biochemistry 20, 849-855 (1979).

Yamamoto, Y., Shimada, S., Nishimura, M.
Purification and molecular properties of 3 polypeptides released from a highly active O_2-evolving photosystem II preparation by Tris-treatment.
FEBS Lett. 151, 49-53 (1983).

Youvan, D.C., Alberti, M., Begusch, H., Bylina, E.J., Hearst, J.E.
Reaction center and light-harvesting I genes from Rhodopseudomonas capsulata.
Proc. Natl. Acad. Sci. USA 81, 189-192 (1984).

Youvan, D.C., Bylina, E.J., Alberti, M., Begusch, H., Hearst, J.E.
Nucleotide and deduced polypeptide sequences of the photosynthetic reaction-center, B870 antenna, and flanking polypeptides from R. capsulata.
Cell 37, 949-957 (1984).

Youvan, D.C., Marrs, B.
Molecular genetics and the light reactions of photosynthesis.
Cell 39,1-3 (1984).

Zielinski, R.E., Price, C.A.
Synthesis of thylakoid membrane proteins by chloroplasts isolated from spinach. Cytochrome b_{559} and P700-chlorophyll a protein.
J. Cell Biol. 85, 435-445 (1980).

Zubay, G.
In vitro synthesis of protein in microbial systems.
Ann. Rev. Genet. 7, 267-287 (1973).

Zurawski, G., Bohnert, H.J., Whitfeld, P.R., Bottomley, W.
Nucleotide sequence of the gene for the M_r 32,000 thylakoid membrane protein from Spinicia oleracea and Nicotiana debneyi predicts a totally conserved primary translation product of M_r 38,950.
Proc. Natl. Acad. Sci. 79, 7699-7703 (1982).

Zurawski, G., Bottomley, W., Whitfeld, P.R.
Structures of the genes for the beta and epsilon subunits of spinach chloroplast ATPase indicate a dicistronic mRNA and an overlapping translation stop/start signal.
Proc. Natl. Acad. Sci. USA 79, 6260-6264 (1982).

Zurawski, G. Perrot, B., Bottomley, W., Whitfeld, P.R.
The structure of the gene for the large subunit of ribulose 1,5-bisphosphate carboxylase from spinach chloroplast DNA.
Nucleic Acids Res. 9, 3251-3270 (1981).

Zusammenfassung

1. Durch DNA-Klonierung und -Sequenzierung sind die Strukturen zweier Gene für chlorophylltragende Proteine des Photosystem II-Reaktionszentrums aus Spinat (Spinacia oleracea) untersucht worden, die des P680 Chlorophyll a Apoproteins, welches die primären photosynthetischen Lichtreaktionen dieses Reaktionszentrums katalysiert, und das 44 kd Chlorophyll a-tragende Protein.

Die Aminosäuresequenz der beider Apoproteine wurden von den DNA Sequenzen abgeleitet. Die größten offenen Leseraster kodieren Proteine der Molekulargewichte 56.246 und 51.785, die 508 bzw. 473 Aminosäurenresten enthalten. Ein Vergleich der beiden Apoproteine über ihre gesamte Länge zeigt 20% Homologie. Dieses legt nahe, daß sie durch Genduplikation entstanden sind.

Beide Proteine besitzen einen hohen Anteil an hydrophoben Aminosäuren. Die Analyse auf Sekundärstruktur nach Kyte und Doolittle (1982) zeigt, daß diese Aminosäuren in sieben Gruppe von je 18 bis 25 Resten auftreten, welche wahrscheinlich in alpha-helikaler Form die Membran durchdringen. Die Transmembransegmente VI und VII beider Proteine sind durch einen besonders großen hydrophilen Bereich von etwa 150 bis 200 Aminosäuren voneinander getrennt. Die durchgehenden Segmente III, IV, V und VII des P680 Chlorophyll a Apoproteins und die Segmente III, V und VII des 44 kd Chlorophyll a Apoproteins haben je ein charakteristisches Paar Histidinreste im Abstand von 13 bzw. 14 Aminosäuren. Diese Reste könnten an der Bindung der Chlorophyll a Molekule beteiligt sein. Spezifizische funktionelle Rollen könnten auch für hoch konservierten Arginin-, Tryptophan-, und Cysteinreste erwartet werden.

2. Es wird angenommen, daß Triazinherbizide den Elektronentransport durch die Bindung an die Chinonbindungsstelle des sogenannten 32 kd Proteins auf der reduzierenden Seite des Photosystems II behindern. Das Gen von einem triazinresistenten Stamm von Amaranthus retroflexus, das dieses Protein verschlüsselt, ist kloniert und sequenziert worden. Das Strukturgen enthält eine einzige Transversion von A nach G . Diese Mutation

verursacht die Veränderung eines Serinrestes nach Glycin bei Position 264 der Aminosäuresequenz. Da ein Serinrest in dieser Position des Proteins in fünf verschiedenen Pflanzenarten vorhanden ist und eine ähnliche Veränderung nach Glycin in einem triazinresistenten Stamm von Amaranthus hybridus gefunden wurde (Hirschberg und McIntosh, 1983), ist die Mutation sehr wahrscheinlich für die Herbizidresistenz verantwortlich.

3. Am 3' Ende des Gens für das herbicidbindende 32 kd-Protein wurde ein Gen für die chloroplastenspezifische $tRNA^{HIS}$ sequenziert.

4. Prokaryotische Transkriptionssignale, wie Promotor- und Terminatorsequenzen, sind in den 5'- bzw. 3'-flankierenden Sequenzen des Gens für das P680 Chlorophyll a Apoproteins lokalisiert. Sieben diskrete RNA-Arten, variierend zwischen 2.0 und mehr als 5.0 Kilobasen, sind zu den kodierenden Sequenzen des Apoproteins komplementär. Dies bedeutet, daß dieser Bereich des plastidären Chromosoms polycistronisch transkribiert werden kann. Die größeren Transkripte enthalten unter anderen auch noch Information für das Cytochrom b_6 und die Untereinheit IV des Cytochrom b/f Komplexes der Thylakoidmembran.

VITA

Julia Elizabeth Bryant Morris was born in St. Louis, Missouri on February 18, 1953 the daughter of the aeronautical engineer William Wilson Bryant and his wife, Phyllis Cattell Bryant. She graduated 44 out of 1219 students from McCluer Senior High School in January of 1971. In the spring of 1971 she attended a semester of college at St. Louis University. In the fall of 1971, she began studies at Washington University and graduated from that institution with an A.B. degree in the spring of 1975. Her major fields were psychology and biology. From July of 1975 through June of 1976, she was employed by Dr. James Nelson, a neuropathologist at Washington University School of Medicine; she recieved training from Dr. Marie Grieder of the Pathology Dept. in electron microscopy. From August of 1976 to September of 1978 she studied at the Graduate School of the University of Texas in Houston and was awarded the degree of Master of Science in Molecular Biology on September 8, 1978. During her studies she was supported by the State of Texas and by a grant from the Rosalie B. Hite Foundation. She married Tom Morris on November 11, 1978. She entered the Universität Düssseldorf in the summer semester of 1980 and recieved the Diplom-Biologe on October 24, 1980. Financial support during the doctorate work was provided by the Graduiertenförderung, Deutsche Forschungs Gemeinschaft, and Bayer-AG, Leverkusen.

FOTODRUCK J. MAINZ GmbH
Neupforte 13, Nähe Markt, D-5100 Aachen
Telefon: (0241) 27305 und (0241) 38606